ELEMENTARY

TEXT-BOOK OF PHYSICS.

BY

PROF. WM. A. ANTHONY, AND PROF. CYRUS F. BRACKETT,

Of Cornell University. *Of Princeton University.*

REVISED BY

PROF. WILLIAM FRANCIS MAGIE,

Of Princeton University.

EIGHTH EDITION, REVISED.

FIRST THOUSAND.

NEW YORK:

JOHN WILEY & SONS.

LONDON: CHAPMAN & HALL, LIMITED.

1897.

Sage

ROBERT DRUMMOND, ELECTROTYPER AND PRINTER, NEW YORK.

PREFACE.

THE design of the authors in the preparation of this work has been to present the fundamental principles of Physics, the experimental basis upon which they rest, and, so far as possible, the methods by which they have been established. Illustrations of these principles by detailed descriptions of special methods of experimentation and of devices necessary for their applications in the arts have been purposely omitted. The authors believe that such illustrations should be left to the lecturer, who, in the performance of his duty, will naturally be guided by considerations respecting the wants of his classes and the resources of his cabinet.

Pictorial representations of apparatus, which can seldom be employed with advantage unless accompanied with full and exact descriptions, have been discarded, and only such simple diagrams have been introduced into the text as seem suited to aid in the demonstrations. By adhering to this plan greater economy of space has been secured than would otherwise have been possible, and thus the work has been kept within reasonable limits.

A few demonstrations have been given which are not usually found in elementary text-books except those which are much more extended in their scope than the present work. This has been done in every case in order that the argument to which the demonstration pertains may be complete, and that the student may be convinced of its validity.

In the discussions the method of limits has been recognized wherever it is naturally involved; the special methods of the cal-

culus, however, have not been employed, since, in most institutions in this country, the study of Physics is commenced before the student is sufficiently familiar with them.

The authors desire to acknowledge their obligations to Wm. F. Magie, Assistant Professor of Physics in the College of New Jersey, who has prepared a large portion of the manuscript and has aided in the final revision of all of it, as well as in reading the proof-sheets.

W. A. ANTHONY,
C. F. BRACKETT.

September, 1887.

REVISER'S PREFACE.

By the courtesy of the authors and publishers of this book, I have been given an opportunity to make a rather extensive revision of it. The principal changes which have been made, besides such slight corrections or supplementary statements as seemed necessary, are, an entire rearrangement and enlargement of the mechanics, and the addition of a discussion of the kinetic theory of matter and of a treatment of magnetism and electricity by the method of tubes of force. The omissions have been largely of statements that would naturally be made by the lecturer or of demonstrations in which the results reached did not warrant the expenditure of time and trouble necessary to master them. I trust that I have adhered throughout to the original design of the authors.

During the last few years I have been using with my classes Selby's "Elementary Mechanics of Solids and Fluids," and have availed myself in many places in the present revision of the suggestions which I received from that admirable book. The additions to the Magnetism and Electricity are based upon the treatment of the subject by J. J. Thomson in his "Elementary Theory of Electricity aud Magnetism."

W. F. Magie.

Princeton University,
February, 1897.

CONTENTS.

LIGHT.

TABLES.

INTRODUCTION.

1. Divisions of Natural Science.—Everything which can affect our senses we call *matter*. Any limited portion of matter, however great or small, is called a *body*. All bodies, together with their unceasing changes, constitute *Nature*.

Natural Science makes us acquainted with the properties of bodies, and with the changes, or *phenomena*, which result from their mutual actions. It is therefore conveniently divided into two principal sections,—*Natural History* and *Natural Philosophy*.

The former describes natural objects, classifies them according to their resemblances, and, by the aid of Natural Philosophy, points out the laws of their production and development. The latter is concerned with the laws which are exhibited in the mutual action of bodies on each other.

These mutual actions are of two kinds: those which leave the essential properties of bodies unaltered, and those which effect a complete change of properties, resulting in loss of identity. Changes of the first kind are called *physical* changes; those of the second kind are called *chemical* changes. Natural Philosophy has, therefore, two subdivisions,—*Physics* and *Chemistry*.

Physics deals with all those phenomena of matter which are not directly related to chemical changes. *Astronomy* is thus a branch of Physics, yet it is usually excluded from works like the present on account of its special character.

It is not possible, however, to draw sharp lines of demarcation between the various departments of Natural Science, for the suc-

cessful pursuit of knowledge in any one of them requires some acquaintance with the others.

2. Methods.—The ultimate basis of all our knowledge of Nature is experience,—experience resulting from the action of bodies on our senses, and the consequent affections of our minds.

When a natural phenomenon arrests our attention, we call the result an *observation.* Simple observations of natural phenomena only in rare instances can lead to such complete knowledge as will suffice for a full understanding of them. An observation is the more complete, the more fully we apprehend the attending circumstances. We are generally not certain that all the circumstances which we note are *conditions* on which the phenomenon, in a given case, depends. In such cases we modify or suppress one of the circumstances, and observe the effect on the phenomenon. If we find a corresponding modification or failure with respect to the phenomenon, we conclude that the circumstance, so modified, is a condition. We may proceed in the same way with each of the remaining circumstances, leaving all unchanged except the single one purposely modified at each trial, and always observing the effect of the modification. We thus determine the conditions on which the phenomenon depends. In other words, we bring *experiment* to our aid in distinguishing between the real conditions on which a phenomenon depends, and the merely accidental circumstances which may attend it.

But this is not the only use of experiment. By its aid we may frequently modify some of the conditions, known to be conditions, in such ways that the phenomenon is not arrested, but so altered in the rate with which its details pass before us that they may be easily observed. Experiment also often leads to new phenomena, and to a knowledge of activities before unobserved. Indeed, by far the greater part of our knowledge of natural phenomena has been acquired by means of experiment. To be of value, experiments must be conducted with system, and so as to trace out the whole course of the phenomenon.

Having acquired our facts by observation and experiment, we

seek to find out how they are related; that is, to discover the *laws* which connect them. The process of reasoning by which we discover such laws is called *induction*. As we can seldom be sure that we have apprehended all the related facts, it is clear that our inductions must generally be incomplete. Hence it follows that conclusions reached in this way are at best only probable; yet their probability becomes very great when we can discover no outstanding fact, and especially so when, regarded provisionally as true, they enable us to predict phenomena before unknown.

In conducting our experiments, and in reasoning upon them, we are often guided by suppositions suggested by previous experience. If the course of our experiment be in accordance with our supposition, there is, so far, a presumption in its favor. So, too, in reference to our reasonings: if all our facts are seen to be consistent with some supposition not unlikely in itself, we say it thereby becomes probable. The term *hypothesis* is usually employed instead of supposition.

Concerning the ultimate modes of existence or action, we know nothing whatever; hence, a law of nature cannot be demonstrated in the sense that a mathematical truth is demonstrated. Yet so great is the constancy of uniform sequence with which phenomena occur in accordance with the laws which we discover, that we have no doubt respecting their validity.

When we would refer a series of ascertained laws to some common agency, we employ the term *theory*. Thus we find in the "wave theory" of light, based on the hypothesis of a universal ether of extreme elasticity, satisfactory explanations of the laws of reflection, refraction, diffraction, polarization, etc.

3. Measurements.—All the phenomena of Nature occur in *matter*, and are presented to us in *time* and *space*.

Time and space are fundamental conceptions: they do not admit of definition. Matter is equally indefinable: its distinctive characteristic is its persistence in whatever state of rest or motion it may happen to have, and the resistance which it offers to any

attempt to change that state. This property is called *inertia.* It must be carefully distinguished from inactivity.

Another essential property of matter is *impenetrability,* or the property of occupying space to the exclusion of other matter.

We are almost constantly obliged, in physical science, to measure the quantities with which we deal. We measure a quantity when we compare it with some standard of the same kind. A simple number expresses the result of the comparison.

If we adopt arbitrary units of length, time, and mass (or quantity of matter), we can express the measure of all other quantities in terms of these so-called *fundamental units.* A unit of any other quantity, thus expressed, is called a *derived unit.*

It is convenient, in defining the measure of derived units, to speak of the ratio between, or the product of, two dissimilar quantities, such as space and time. This must always be understood to mean the ratio between, or the product of, the numbers expressing those quantities in the fundamental units. The result of taking such a ratio or product of two dissimilar quantities is a number expressing a third quantity in terms of a derived unit.

4. **Unit of Length.**—The *unit of length* usually adopted in scientific work is the *centimetre.* It is the one hundredth part of the length of a certain piece of platinum, declared to be a standard by legislative act, and preserved in the archives of France. This standard, called the *metre,* was designed to be equal in length to one ten-millionth of the earth's quadrant.

The operation of comparing a length with the standard is often difficult of direct accomplishment. This may arise from the minuteness of the object or distance to be measured, from the distant point at which the measurement is to end being inaccessible, or from the difficulty of accurately dividing our standard into very small fractional parts. In all such cases we have recourse to indirect methods, by which the difficulties are more or less completely obviated.

The *vernier* enables us to estimate small fractions of the unit of length with great convenience and accuracy. It consists of an

accessory piece, fitted to slide on the principal scale of the instrument to which it is applied. A portion of the accessory piece, equal to n minus one or n plus one divisions of the principal scale, is divided into n divisions. In the former case, the divisions are numbered in the same sense as those of the principal scale; in the latter, they are numbered in the opposite sense. In either case we can measure a quantity accurately to the one nth part of one of the primary divisions of the principal scale. Fig. 1 will make the construction and use of the vernier plain.

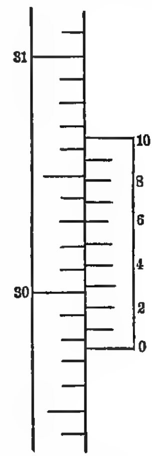

FIG. 1.

In Fig. 1, let 0, 1, 2, 3 . . . 10 be the divisions on the vernier; let 0, 1, 2, 3 . . . 10 be any set of consecutive divisions on the principal scale.

If we suppose the 0 of the vernier to be in coincidence with the limiting point of the magnitude to be measured, it is clear that, from the position shown in the figure, we have 29.7, expressing that magnitude to the nearest tenth; and since the sixth division of the vernier coincides with a whole division of the principal scale, we have $\frac{6}{10}$ of $\frac{1}{10}$, or $\frac{6}{100}$, of a principal division to be added: hence the whole value is 29.76.

The *micrometer screw* is also much employed. It consists of a carefully cut screw, accurately fitting in a nut. The head of the screw carries a graduated circle, which can turn past a fixed line. This is frequently the straight edge of a scale with divisions equal in magnitude to the pitch of the screw. These divisions will then show through how many revolutions the screw is turned in any given trial; while the divisions on the graduated circle will show the fractional part of a revolution, and consequently the fractional part of the pitch that must be added. If the screw be turned through n revolutions, as shown by the scale, and through an additional fraction, as shown by the divided circle, it will pass through n times the pitch of the screw, and an additional fraction of the pitch determined by the ratio of the number of divisions read

from 0 on the divided circle to the whole number into which it is divided.

The *cathetometer* is used for measuring differences of level. A graduated scale is cut on an upright bar, which can turn about a vertical axis. Over this bar slide two accurately fitting pieces, one of which can be clamped to the bar at any point, and serve as the fixed bearing of a micrometer screw. The screw runs in a nut in the second piece, which has a vernier attached, and carries a horizontal telescope furnished with cross-hairs. The telescope having been made accurately horizontal by means of a delicate level, the cross-hairs are made to cover one of the two points, the difference of level between which is sought, and the reading upon the scale is taken; the fixed piece is then unclamped, and the telescope raised or lowered until the second point is covered by the cross-hairs, and the scale reading is again taken. The difference of scale reading is the difference of level sought.

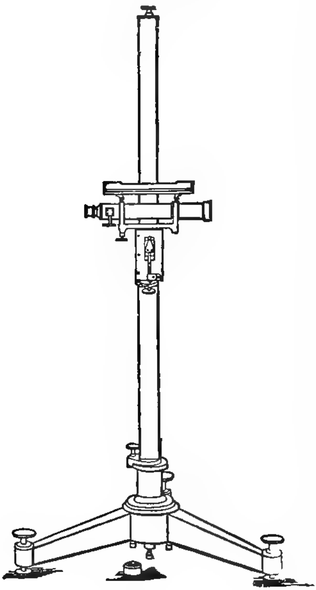

Fig. 2.

The *dividing engine* may be used for dividing scales or for comparing lengths. In its usual form it consists essentially of a long micrometer screw, carrying a table, which slides, with a motion accurately parallel with itself, along fixed guides, resting on a firm support. To this table is fixed an apparatus for making successive cuts upon the object to be graduated.

The object to be graduated is fastened to the fixed support. The table is carried along through any required distance determined by the motion of the screw, and the cuts can be thus made at the proper intervals.

The same instrument, furnished with microscopes and accessories, may be employed for comparing lengths with a standard. It may then be called a *comparator*.

The *spherometer* is a special form of the micrometer screw. As

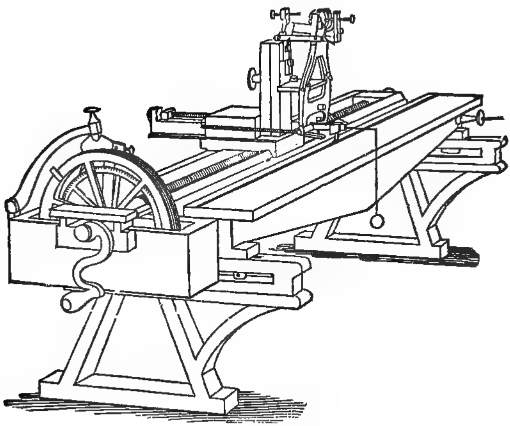

FIG. 3.

its name implies, it is primarily used for measuring the curvature of spherical surfaces.

It consists of a screw with a large head, divided into a great number of parts, turning in a nut supported on three legs terminating in points, which form the vertices of an equilateral triangle. The axis of revolution of the screw is perpendicular to the plane of the triangle, and passes through its centre. The screw ends in a point which may be brought into the same plane with the points of the legs. This is done by placing the legs on a truly plane surface, and turning the screw till its point is just in contact with the surface. The sense of touch will enable one to decide with great nicety when the screw is turned far enough. If, now, we note the reading of the divided scale and also that of the divided head, and then raise the screw, by turning it backward, so that the given curved surface may exactly coincide with the four points, we can compute the radius of curvature from the difference of the two

readings and the known length of the side of the triangle formed by the points of the tripod.

5. Unit of Time.—The *unit of time* is the *mean time second,* which is the $\frac{1}{86400}$ of a mean solar day. We employ the *clock,* regulated by the pendulum or the chronometer balance, to indicate seconds. The clock, while sufficiently accurate for ordinary use, must for exact investigations be frequently corrected by astronomical observations.

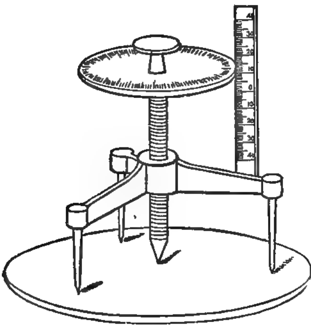

FIG. 4.

Smaller intervals of time than the second are measured by causing some vibrating body, as a tuning-fork, to trace its path along some suitable surface, on which also are recorded the beginning and end of the interval of time to be measured. The number of vibrations traced while the event is occurring determines its duration in known parts of a second.

In estimating the duration of certain phenomena giving rise to light, the revolving mirror may be employed. By its use, with proper accessories, intervals as small as forty billionths of a second have been estimated.

6. Unit of Mass.—The *unit of mass* usually adopted in scientific work is the *gram.* It is equal to the one-thousandth part of a certain piece of platinum, called the *kilogram,* preserved as a standard in the archives of France. This standard was intended to be equal in mass to one cubic decimetre of water at its greatest density.

Masses are compared by means of the *balance,* the construction of which will be discussed hereafter.

7. Measurement of Angles.—Angles are usually measured by reference to a *divided circle* graduated on the system of division upon which the ordinary trigonometrical tables are based. A pointer or an arm turns about the centre of the circle, and the

angle between two of its positions is measured in degrees on the arc of the circle. For greater accuracy, the readings may be made by the help of a vernier. To facilitate the measurement of an angle subtended at the centre of the circle by two distant points, a telescope with cross-hairs is mounted on the movable arm.

In theoretical discussions the unit of angle often adopted is the *radian*, that is, the angle subtended by the arc of a circle equal to its radius. In terms of this unit, a semi-circumference equals $\pi = 3.141592$. The radian, measured in degrees, is $57° 17' 44.8.''$

8. Dimensions of Units.—Any derived unit may be represented by the product of certain powers of the symbols representing the fundamental units of length, mass, and time.

Any equation showing what powers of the fundamental units enter into the expression for the derived unit is called its *dimensional equation.* In a dimensional equation time is represented by T, length by L, and mass by M. To indicate the dimensions of any quantity, the symbol representing that quantity is enclosed in brackets.

For example, the unit of area varies as the square of the unit of length; hence its dimensional equation is $[\text{area}] = L^2$. In like manner, the dimensional equation for volume is $[\text{vol.}] = L^3$.

9. Systems of Units.—The system of units adopted in this book, and generally employed in scientific work, based upon the centimetre, gram, and second, as fundamental units, is called the *centimetre-gram-second system* or the C. G. S. system. A system based upon the foot, grain, and second was formerly much used in England. One based upon the millimetre, milligram, and second is still sometimes used in Germany.

MECHANICS.

CHAPTER I.

MECHANICS OF MASSES.

10. It is an obvious fact of Nature that material bodies move from one place to another, and that their motions are effected at different rates and in different manners. Continued experience has shown that these motions are independent of many of the characteristics of the bodies ; they depend on the arrangement and condition of surrounding bodies, and on the fundamental property of matter, called inertia. The science of *Mechanics* treats of the motions here referred to, and in a wider sense of those phenomena presented by bodies which depend more or less directly upon their masses.

The general subject of Mechanics is usually divided, in extended treatises, into two topics,—*Kinematics* and *Dynamics*. In the first are developed, by purely mathematical methods, the laws of motion considered in the abstract, independent of any causes producing it, and of any substance in which it inheres ; in the second these mathematical relations are extended and applied, by the aid of a few inductions drawn from universal experience, to the explanation of the motions of bodies, and the discussion of the interactions which are the occasion of those motions.

For convenience, the subject of Dynamics is further divided into *Statics*, which treats of forces as maintaining bodies in equilibrium and at rest, and *Kinetics*, which treats of forces as setting bodies in motion.

It has been found more convenient to neglect these formal distinctions in the very brief presentation of the subject which will be given in this book.

11. Configuration and Displacement.—An assemblage of points may be completely described by selecting some one point as a point of reference and assigning to each of the others a definite distance and direction measured from this fixed point. Such a set of points is called a *system* of points, and the assemblage of distances and directions which characterize it is called its *configuration*. The *motion* of one or more of the points is recognized by a change in the configuration. The change in position of any one point, determined by the distance between its initial and final positions and the direction of the line drawn between those positions, is called the *displacement* of the point.

Any particle in the system may be taken as the fixed point of reference, and the motion of the others may be measured from it. Thus, for example, high-water mark on the shore may be taken as the fixed point in determining the rise and fall of the tides; or, the sun may be assumed to be at rest in computing the orbital motions of the planets. We can have no assurance that the particle which we assume as fixed is not really in motion as a part of some larger system; indeed, in almost every case we know that it is thus in motion. As it is impossible to conceive of a point in space recognizable as fixed and determined in position, our measurements of motion must always be relative.

12. Composition and Resolution of Displacements.—If a point undergo two or more successive displacements, the final displacement is obviously given by the line joining its initial to its final position. This displacement is called the *resultant* of the others. If the point considered be referred to a point which is itself displaced relative to a third point taken as fixed, the motion of the moving point relative to the fixed point may be considered as resulting from a combination of the displacement of the first point relative to the second point, and the displacement of the second point relative to the third or fixed point. These simultaneous

displacements are combined as if they were successive displacements. Representing them both by straight lines, of which the length measures the amount of the displacement, and the direction the direction of the displacement (Fig. 5), we apply the initial point of the second of these lines to the final point of the first and join the initial point of the first to the final point of the second. The line thus drawn is the resultant of the simultaneous displacements. The two displacements of which the resultant is thus obtained are called the *components*.

13. Vector Addition and Subtraction.—Any concept which is completely described when its magnitude and direction are given is called a *vector*. The sum of two vectors is the vector equivalent to them both. It is obtained by the rule just given for the composition of two displacements, or by the following equivalent rule: Draw from any point the two straight lines which represent the vectors, and upon them construct a parallelogram; the diagonal of this parallelogram, drawn from the point of origin, is the resultant vector or the vector sum. Thus OC (Fig. 5) is the resultant of OA and OB. This construction is called the parallelogram construction or the *parallelogram law.* If more than two vectors are to be added, the resultant of two of them may be added to the third, the resultant thus obtained to the fourth, and so on until all the vectors have been combined. This addition is more easily made by drawing the vectors in succession, so that they form the sides of a polygon (Fig. 6), the initial point of each vector coinciding with the final point of the one preceding it. In general this polygon is not closed, and the line required to close it, drawn from the initial point of the first vector to the final point of the last, is the sum of the vectors. This construction is called the polygon construction or the *polygon law.*

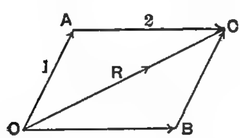

FIG. 5.

The difference of two vectors is the vector which added to one of the two will give the other. It is obtained by drawing from a given point the lines representing the vectors, and drawing a line

from the final point of the subtrahend to the final point of the minuend. This line represents the vector difference of the two vectors. Thus AC (Fig. 5) is the difference between OC and OA. The same difference may be obtained by the following method: If two lines, equal in length, be drawn in opposite directions, they represent two vector quantities which have the same magnitude but are affected with opposite signs. If, therefore, a vector be given which is to be subtracted from another, it may be replaced by a vector of the same magnitude having the opposite direction, and the resultant obtained by adding this vector to the one which serves as the minuend is the difference of the two given vectors.

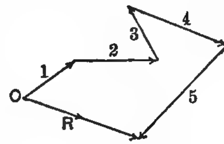

FIG. 6.

14. Resolution and Composition of Vectors.—It is in many cases convenient to obtain component vectors which are equivalent to a given vector. If one component be completely given, the other is obtained by vector subtraction. If two components be desired, and their directions be given so that they and the original vector are in the same plane, their magnitudes may be determined by drawing from a common origin lines of indefinite length in the given directions, drawing from the same origin the line representing the given vector, and drawing from its final point lines parallel to the given directions. The sides of the parallelogram thus constructed represent the component vectors in these given directions.

If three components be desired in three given directions not in the same plane, and so placed that the given vector does not lie in a plane containing any two of these directions, they may be found by constructing upon lines drawn in these directions a parallelepiped of which the diagonal is the given vector. This construction is most frequently used when the three directions are at right angles to one another. Representing the angles between them and the direction of the given vector by α, β, γ, the component vectors are proportional to $\cos\alpha$, $\cos\beta$, $\cos\gamma$. If these three directions be the directions of the axes of a system of rectangular coordinates, these cosines are called the direction cosines of the vector.

The composition of vectors is often conveniently effected by resolving them in this way along the three coordinate axes; their components along each of these axes may then be added algebraically, and the vector obtained by combining the three sums is the required resultant vector. Thus if the vectors R_1 $R_2 \ldots R_n$ be given, making angles with the x, y, z-axes of which the cosines are $\lambda_1, \lambda_2 \ldots \lambda_n$, $\mu_1, \mu_2 \ldots \mu_n$, $\nu_1, \nu_2 \ldots \nu_n$, respectively, the sums of the components of these vectors along the axes are

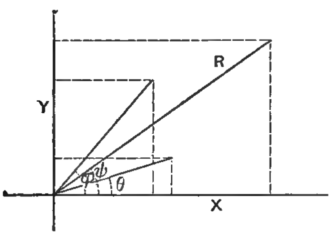

FIG. 7.

$$\left.\begin{aligned} X &= R_1\lambda_1 + R_2\lambda_2 + \ldots + R_n\lambda_n; \\ Y &= R_1\mu_1 + R_2\mu_2 + \ldots + R_n\mu_n; \\ Z &= R_1\nu_1 + R_2\nu_2 + \ldots + R_n\nu_n. \end{aligned}\right\} \quad (1)$$

The resultant vector is

$$R = \sqrt{X^2 + Y^2 + Z^2},$$

and its direction cosines are

$$\frac{X}{R}, \frac{Y}{R}, \frac{Z}{R},$$

respectively.

When only two vectors are given, they may be resolved along two axes in the plane of the vectors. In this case, if the angles made by the vectors R_1, R_2 with the x-axis be ϕ, ϑ, respectively, (Fig. 7,) the component sums are

$$\left.\begin{aligned} X &= R_1 \cos\phi + R_2 \cos\vartheta, \\ Y &= R_1 \sin\phi + R_2 \sin\vartheta. \end{aligned}\right\} \quad (2)$$

The resultant vector is $R = \sqrt{X^2 + Y^2}$, and the angle ψ which it makes with the x-axis is given by $\cos\psi = \dfrac{X}{R}$ or $\tan\psi = \dfrac{Y}{X}$.

15. Description of Motion.—If we observe a system of points in motion, we perceive not only the displacements of the points, but also that these displacements are in some way connected with the time required for their accomplishment. If we know the law of this connection, we may describe the motion at any desired instant, by the aid of certain derived concepts, which are now to be studied.

If a variable quantity be a function of the time, it is usual in

Mechanics to call the limit of the ratio of a small change in that quantity to the time-interval in which it occurs the *rate* of change of the quantity. This ratio is the differential coefficient of the quantity with respect to time. Other differential coefficients which occur in Mechanics, in which the independent variable is not the time, are sometimes spoken of as rates, though not frequently. The motion of a point is described when we know not only the *path* along which it is displaced, but the rates connected with its displacement.

16. Velocity.—The rate of displacement of a point is called its *velocity*. If the point move in a straight line, and describe equal spaces in any arbitrary equal times, its velocity is *constant*. A constant velocity is measured by the ratio of the space traversed by the point to the time occupied in traversing that space. If s_0 and s represent the distances of the point from a fixed point on its path at the instants t_0 and t, then its velocity is represented by

$$v = \frac{s - s_0}{t - t_0}. \tag{3}$$

If the path of the point be curved, or if the spaces described by the point in equal times be not equal, its velocity is *variable*. The path of a point moving with a variable velocity may be approximately represented by a succession of very small straight lines, which, if the real path be curved, will differ in direction, along which the point moves with constant velocities which may differ in amount. The velocity in any one of these straight lines is represented by the formula $v = \frac{s - s_0}{t - t_0}$. As the interval of time $t - t_0$ approaches zero, each of the spaces $s - s_0$ will become indefinitely small, and in the limit the imaginary path will coincide with the real path. The limit of the expression $\frac{s - s_0}{t - t_0}$ will represent the velocity of the point along the tangent to the path at the time $t = t_0$, or, as it is called, the velocity in the path. This limit is usually expressed by $\frac{ds}{dt}$.

The practical unit of velocity is the velocity of a point moving uniformly through one centimetre in one second.

The dimensions of velocity are LT^{-1}.

Velocity, which is fully defined when its magnitude and direction are given, is a vector quantity, and may be represented by a straight line. Velocities may therefore be compounded and resolved by the rules already given for the composition and resolution of vectors.

17. Acceleration.—When the velocity of a point varies, either by a change in its magnitude, or by a change in its direction, or by changes in both, the rate of change is called the *acceleration* of the point. Acceleration is either positive or negative, according as the velocity increases or diminishes. If the path of the point be a straight line, and if equal changes in velocity occur in equal times, its acceleration is *constant*. It is measured by the ratio of the change in velocity to the time during which that change occurs. If v_0 and v represent the velocities of the point at the instants t_0 and t, then its acceleration is represented by

$$a = \frac{v - v_0}{t - t_0}. \tag{4}$$

If the path of the point be curved, or if the changes in velocity in equal times be not equal, the acceleration is *variable*. A variable acceleration in a curved path may always be resolved into two components, one of which is tangent and the other normal to the path. We will consider the case in which the path lies in a plane.

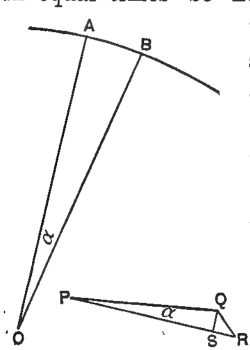

Fig. 8.

Let A and B (Fig. 8) be two points in the path very near each other, from which normals are drawn on the concave side of the curve, meeting at the point O, and making with each other the angle α. In the limit, as α vanishes, the lines OA and OB become equal and are radii of curvature of the path at the

point A. Draw the lines PQ and PR in the directions of the tangents at A and B, equal to the velocities v_0 and v of the point at A and B respectively. The line QR is the change in the velocity of the point during the time in which it traverses the distance AB. Draw the line QS perpendicular to PR. The angle QPR, being the angle between the tangents at A and B, equals the angle α. In the limit, as α vanishes, v and v_0 differ by the infinitesimal SR, and QS equals $v\alpha$. The line SR represents the change in the numerical magnitude of the velocity during the time $t - t_0$, and the rate of that change, which takes place along the tangent to the path, is given by

$$a_t = \frac{v - v_0}{t - t_0}. \tag{5}$$

The line QS represents the change in velocity during the same time along the normal to the path. The acceleration along that normal is therefore $\frac{v\alpha}{t - t_0}$. Now under the conditions assumed in these statements $AB = r\alpha$, and $\frac{AB}{t - t_0} = v$, the velocity of the point. Hence $v = \frac{r\alpha}{t - t_0}$, and the acceleration along the normal to the path is

$$a_n = \frac{v^2}{r}. \tag{6}$$

If the path be a straight line, the normal acceleration vanishes, and the whole acceleration is given by the limit of the ratio $\frac{v - v_0}{t - t_0} = \frac{dv}{dt}$. If the path be a circle, and if the point move in it uniformly, the whole acceleration is given by $\frac{v^2}{r}$.

The unit of acceleration is that of a point, the velocity of which changes at a uniform rate by one unit of velocity in one second.

The dimensions of acceleration are LT^{-2}.

Acceleration is completely described when its magnitude and

direction are given. It is therefore a vector quantity and may be represented by a straight line. Two or more accelerations may be compounded by the rules for the composition of vectors.

18. Angular Velocity and Acceleration.—The angle contained by the line passing through two points, one of which is in motion, and any assumed line passing through the fixed point, will, in general, vary. The rate of its change is called the *angular velocity* of the moving point. If ϕ and ϕ_0 represent the angles made by the moving line with the fixed line at the instants t and t_0, then the angular velocity, if constant, is measured by

$$\omega = \frac{\phi - \phi_0}{t - t_0}. \tag{7}$$

If variable, it is measured by the limit of the same expression, $\frac{d\phi}{dt} = \frac{\phi - \phi_0}{t - t_0}$, as the interval $t - t_0$ becomes indefinitely small.

The *angular acceleration* is the rate of change of angular velocity. If constant, it is measured by

$$\alpha = \frac{\omega - \omega_0}{t - t_0}. \tag{8}$$

If variable, it is measured by the limit of the same expression, $\frac{d\omega}{dt} = \frac{\omega - \omega_0}{t - t_0}$, as the interval $t - t_0$ becomes indefinitely small.

If the radian be taken as the unit of angle, the dimensions of angle become $\left[\frac{\text{arc}}{\text{radius}}\right] = \frac{L}{L} = 1$. Hence the dimensions of angular velocity are T^{-1}, and of angular acceleration T^{-2}.

If any point be revolving about a fixed point as a centre, its velocity in the circle is equal to the product of its angular velocity and the length of the radius of the circle.

19. Linear Motion with Constant Acceleration. — The space $s - s_0$ traversed by a point moving with a constant acceleration a, during a time $t - t_0$, is determined by considering that, since the acceleration is constant, the average velocity $\frac{v + v_0}{2}$ for the time

$t - t_0$, multiplied by $t - t_0$, will represent the space traversed; hence

$$s - s_0 = \frac{v + v_0}{2}(t - t_0); \tag{9}$$

or, since $\frac{v}{2} = \frac{v_0 + a(t - t_0)}{2}$, we have, in another form,

$$s - s_0 = v_0(t - t_0) + \tfrac{1}{2}a(t - t_0)^2. \tag{9a}$$

Multiplying equations (4) and (9), we obtain

$$v^2 = v_0^2 + 2a(s - s_0). \tag{10}$$

When the point starts from rest, $v_0 = 0$; and if we take the starting-point as the origin from which to reckon s, and the time of starting as the origin of time, then $s_0 = 0$, $t_0 = 0$, and equations (4), (9a), and (10) become $v = at$, $s = \frac{1}{2}at^2$, and $v^2 = 2as$.

Formula (9a) may also be obtained by a geometrical construction. At the extremities of a line AB (Fig. 9), equal in length to $t - t_0$, erect perpendiculars AC and BD, proportional to the initial and final velocities of the moving point. For any interval of time Aa, so short that the velocity during it may be considered constant, the space described is represented by the rectangle Ca, and the space described in the whole time $t - t_0$, by a point moving with a velocity increasing by successive equal increments, is represented by a series of rectangles, eb, fc, gd, etc., described on equal bases, ab, bc, cd, etc.

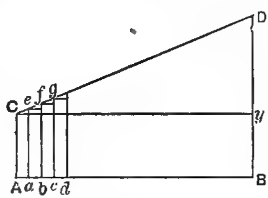

FIG. 9.

If ab, bc . . . be diminished indefinitely, the sum of the areas of the rectangles can be made to approach as nearly as we please the area of the quadrilateral $ABCD$. This area, therefore, represents the space traversed by the point, having the initial velocity v_0, and moving with the acceleration a during the time $t - t_0$. But $ABCD$ is equal to $AC(t - t_0) + (BD - AC)(t - t_0) \div 2$; whence

$$s - s_0 = v_0(t - t_0) + \tfrac{1}{2}a(t - t_0)^2. \tag{9a}$$

20. Angular Motion with Constant Angular Acceleration.—If a

point move in a circle its velocity is equal to the product of its angular velocity and the radius of the circle; its acceleration in the circle is equal to the product of its angular acceleration and the radius of the circle. If its angular acceleration be constant, the relations between the distance traversed by it in the circle, its velocity, its acceleration in the circle and the time are the same as those expressed in equations (9), (9a), (10). Substituting for these quantities their equivalents in terms of the angular magnitudes involved, we obtain the following relations among these angular magnitudes:

$$\phi - \phi_0 = \frac{\omega + \omega_0}{2}(t - t_0); \tag{11}$$

$$\phi - \phi_0 = \omega_0(t - t_0) + \tfrac{1}{2}\alpha(t - t_0)^2; \tag{12}$$

$$\omega^2 = \omega_0^2 + 2\alpha(\phi - \phi_0). \tag{13}$$

If the line describing the angle start from rest, $\omega_0 = 0$, and if we take the line in this position as the initial line from which to reckon ϕ, and the time of starting as the origin of time, then $\phi_0 = 0$, $t_0 = 0$, and equations (8), (12), (13), become $\omega = \alpha t$, $\phi = \frac{1}{2}\alpha t^2$, and $\omega^2 = 2\alpha\phi$.

21. Simple Harmonic Motion.—If a point move in a circle with a constant velocity, the point of intersection of a diameter and a perpendicular drawn from the moving point to this diameter will have a *simple harmonic motion.* Its velocity at any instant will be the projection of the velocity of the point moving in the circle at that instant upon the diameter. The radius of the circle is the *amplitude* of the motion. The *period* is the time between any two successive recurrences of a particular condition of the moving point. The position of a point executing a simple harmonic motion can be expressed in terms of the interval of time which has elapsed since the point last passed through the middle of its path in the positive direction. This interval of time, when expressed as a fraction of the period, is the *phase.*

We further define rotation in the positive direction as that rotation in the circle which is contrary to the motion of the hands of

a clock, or counter-clockwise. Motion from left to right in the diameter is also considered positive. Displacement to the right of the centre is positive, and to the left negative.

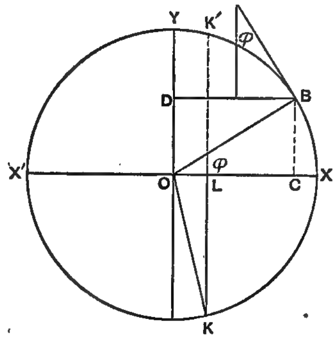

FIG. 10.

If a point start from X (Fig. 10), the position of greatest positive *elongation*, with a simple harmonic motion, its distance s from O or its *displacement* at the end of the time t, during which the point in the circle has moved through the arc BX, is $OC = OB \cos \phi$. Now, OB is equal to OX, the amplitude, represented by a. If ω represent the angular velocity of the moving point, we have $\phi = \omega t$. Hence we have

$$s = a \cos \omega t. \tag{14}$$

To find the velocity at the point C, we must resolve the velocity of the point moving in the circle into its components parallel to the axes. The component at the point C along OX is $V \sin \phi$; or, since $V = \omega a$,

$$v = - \omega a \sin \omega t, \tag{15}$$

remembering that motion from right to left is considered negative.

The acceleration at the point C is the component along OX of the acceleration of the point moving in the circle. The acceleration of B is $-\frac{v^2}{a}$, the minus sign being given because this acceleration is directed opposite to the positive direction of the radius. The component at C along OX is

$$f = -\frac{v^2}{a} \cos \omega t \text{ or } f = -\omega^2 a \cos \omega t = -\omega^2 s. \tag{16}$$

This formula shows that the acceleration in a simple harmonic

motion is proportional to the displacement, and that acceleration to the right of O is negative, to the left of O positive.

In these formulas the angular velocity ω may be replaced by an equivalent factor involving the period T. For, the line drawn from O to the point moving in the circle sweeps out the angle 2π in the time T, so that $\omega = \frac{2\pi}{T}$.

It is often convenient to reckon time from some other position than that of greatest positive elongation. In that case the time required for the moving point to reach its greatest positive elongation from that position, or the angle described by the corresponding point in the circumference in that time, is called the *epoch* of the new starting-point. In determining the epoch, it is necessary to consider, not only the position, but the direction of motion, of the moving point at the instant from which time is reckoned. Thus, if L, corresponding to K in the circumference, be taken as the starting-point, the epoch is the time required to describe the path LX. But if L correspond to the point K' in the circumference, the motion in the diameter is negative, and the epoch is the time required for the moving-point to go from L through O to X' and back to X.

The epochs in the two cases, expressed in angle, are, in the first, the angle measured by the arc KX; and, in the second, the angle measured by the arc $K'X'KX$.

Choosing K in the circle, or L in the diameter, as the point from which time is to be reckoned, the angle ϕ equals angle KOB − angle KOX, or $\omega t - \epsilon$, where t is now the time required for the moving point to describe the arc KB, and ϵ is the epoch, or the angle KOX.

The formulas then become

$$s = a \cos(\omega t - \epsilon);$$

$$v = -\omega a \sin(\omega t - \epsilon);$$

$$f = -\omega^2 a \cos(\omega t - \epsilon);$$

Returning to our first suppositions, letting X be the point from which epoch and time are reckoned, it is plain that, since

$$BC = a \sin \phi = a \cos\left(\phi - \frac{\pi}{2}\right) = a \cos\left(\omega t - \frac{\pi}{2}\right),$$

the projection of B on the diameter OY also has a simple harmonic motion, differing in epoch from that in the diameter OX by $\frac{\pi}{2}$. It follows immediately that the composition of two simple harmonic motions at right angles to each other, having the same amplitude and the same period, and differing in epoch by a right angle, will produce a motion in a circle of radius a with a constant velocity. More generally, the coordinates of a point moving with two simple harmonic motions at right angles to one another are

$$x = a \cos(\phi - \epsilon) \quad \text{and} \quad y = b \cos \phi'.$$

If ϕ and ϕ' are commensurable, that is, if $\phi' = n\phi$, the curve is re-entrant. Making this supposition,

$$x = a \cos \phi \cos \epsilon + a \sin \phi \sin \epsilon, \quad \text{and} \quad y = b \cos n\phi.$$

Various values may be assigned to a, to b, and to n. Let a equal b and n equal 1; then

$$x = y \cos \epsilon + (a^2 - y^2)^{\frac{1}{2}} \sin \epsilon;$$

from which

$$x^2 - 2xy \cos \epsilon + y^2 \cos^2 \epsilon = a^2 \sin^2 \epsilon - y^2 \sin^2 \epsilon,$$

or,

$$x^2 - 2xy \cos \epsilon + y^2 = a^2 \sin^2 \epsilon.$$

This becomes, when $\epsilon = 90°$, $x^2 + y^2 = a^2$, the equation for a circle. When $\epsilon = 0°$, it becomes $x - y = 0$, the equation for a straight line through the origin, making an angle of 45° with the axis of X. With intermediate values of ϵ, it is the equation for an ellipse. If we make $n = \frac{1}{2}$, we obtain, as special cases of the curve, a parabola and a lemniscate, according as $\epsilon = 0°$ or $90°$. If a and b are unequal, and $n = 1$, we get, in general, an ellipse.

We shall now show, in the simplest case, the result of compounding two simple harmonic motions which differ only in epoch

and are in the same line. Let their displacements be represented by $s = a\cos\omega t$, and $s' = a\cos(\omega t - \epsilon)$.

The resultant displacement is the sum of the displacements due to each; hence

$$\begin{aligned} s + s' &= a[\cos\omega t + \cos(\omega t - \epsilon)], \\ &= a[\cos\omega t + \cos\omega t\cos\epsilon + \sin\omega t\sin\epsilon], \\ &= a[\cos\omega t\,(1 + \cos\epsilon) + \sin\omega t\sin\epsilon]. \end{aligned}$$

If for brevity we assume a value A and an angle ϕ such that $A\cos\phi = a(1 + \cos\epsilon)$, and $A\sin\phi = a\sin\epsilon$, we may represent the last value of $s + s'$ by $A\cos(\omega t - \phi)$. From the two equations containing A, we obtain, by adding the squares of the values of $A\sin\phi$ and $A\cos\phi$, $A = (2a^2 + 2a^2\cos\epsilon)^{\frac{1}{2}}$; and, by dividing the value of $A\sin\phi$ by that of $A\cos\phi$, we obtain $\phi = \tan^{-1}\dfrac{\sin\epsilon}{1 + \cos\epsilon}$. The displacement thus becomes

$$s + s' = a(2 + 2\cos\epsilon)^{\frac{1}{2}}\cos\left(\omega t - \tan^{-1}\frac{\sin\epsilon}{1 + \cos\epsilon}\right). \qquad (17)$$

This equation is of great value in the discussion of problems in optics.

The principle suggested by the result of the above discussion, that the resultant of the composition of two simple harmonic motions is a periodic motion of which the elements depend on those of the components, can be easily seen to hold generally.

A very important theorem, of which this principle is the converse, was given by Fourier. It may be stated as follows: Any complex periodic function may be resolved into a number of simple harmonic functions of which the periods are commensurable with that of the original function.

22. Force.—When we lift or sustain a weight, stretch a spring, or throw a ball, we are conscious of a muscular effort which we designate as a *force*. Since no change can be perceived in the weight if it be suspended from a cord, or in the spring if it be held stretched by being fastened to a hook, and since the ball moves in just the same way if it be projected from a gun, we conclude that

bodies can exert force on one another. This conclusion is not strictly justifiable, and our comparison of the action of one body on another to the action of our muscles may be only a convenient analogy.

If we throw a weight by exerting a certain effort for a short time, and then by exerting an equal effort for a longer time, we find that the velocity acquired by the weight is greater in the latter case. If we apply different efforts for the same time in throwing the same weight, we find that the effort which we are conscious of as greater gives the weight a greater velocity than that effort which we are conscious of as less. We may substitute for the forces exerted by our muscles those forces which we have assumed by analogy to act between bodies. Relying upon the uniformity with which these forces act, as determined by universal experience, we can exhibit, more precisely than by the use of our muscular effort, the relations which obtain between the force exerted and the motion caused by it. As our experiments increase in precision, and as one disturbing cause after another is eliminated, we find that the velocity acquired by a given body acted on by a given force increases in proportion to the time during which the force acts, or, as may be said, a constant force produces a uniform acceleration. Further, if different forces act on the same body for the same time, the velocities produced are proportional to the forces. If F represent the magnitude of the force, t the time during which it acts, v the velocity which the body acquires, and m a proportional factor, the results of these experiments may be embodied in the formula

$$Ft = mv. \tag{18}$$

The factor m is called the *mass* or the *inertia* of the body. Since $\frac{v}{t}$ measures the acceleration of the body, this equation is equivalent to

$$F = ma. \tag{19}$$

The dimensions of force are MLT^{-2}.

The practical unit of force is the *dyne,* which is the force that

can impart to a gram of matter one unit of acceleration; that is to say, one unit of velocity in one second.

23. Impulse.—The product Ft is called the *impulse.* If the force which acts upon the body vary during the time, the impulse is determined by dividing the time into intervals so small that the force which acts during any one of them may be considered constant, forming the product Ft for each interval, and adding those products.

24. Momentum.—The product mv is called the *momentum* of the body. It is sometimes defined as the quantity of motion of the body; in Newton's laws, which follow, the word "motion" is equivalent to momentum, when it designates a measurable quantity.

25. Laws of Motion.—The relation between force and acceleration, which is embodied in the formula $F = ma$, was first perceived by Galileo, and illustrated by him by the laws of falling bodies. This relation may be expressed otherwise by the statement that the effect of a force on a body is independent of the motion of the body. Newton, who first formulated the fundamental facts of motion in such a form that they can be made the basis of a science of Mechanics, extended Galileo's principle by recognizing that when several forces act on a body at once the effect of each is independent of the others. Newton's Laws of Motion, in which the fundamental facts of motion are stated, are as follows:

LAW I.—Every body continues in its state of rest or of uniform motion in a straight line, except in so far as it may be compelled by external forces to change that state.

LAW II.—Change of motion is proportional to the external force applied, and takes place in the direction of the straight line in which the force acts.

LAW III.—To every action there is always an equal and contrary reaction; or, the mutual actions of any two bodies are always equal and oppositely directed.

These laws cannot be applied, without some limitations or modi-

fications, to all bodies. They are to be understood as applying to very small masses, for which we can neglect the velocity of rotation in comparison with the velocity of translation. Such a mass is called a *particle*. A particle may also be defined as a mass concentrated at a point. Another definition will be given in § 37.

These laws of motion are not immediately susceptible of proof; they are abstractions, which can be illustrated but not proved by experiment. They cannot be referred to any more ultimate principles deduced from our observation of Nature, and are therefore to be considered as postulates upon which the science of Mechanics is erected. The question of their validity as expressions of the mode of motion of matter is one which lies outside the range of the purely physical study of the subject.

26. Discussion of the Laws of Motion.—(1) The first law is a statement of the important truths, that motion, as well as rest, is a natural state of matter; that moving bodies, when entirely free to move, proceed in straight lines, and describe equal spaces in equal times; and that any deviation from this uniform rectilinear motion is caused by a force.

That a body at rest should continue indefinitely in that state seems perfectly obvious as soon as the proposition is entertained; but that a body in motion should continue to move in a straight line is not so obvious, since motions with which we are familiar are frequently arrested or altered by causes not at once apparent. This important truth, which is forced upon us by observation and experience, may, however, be presented so as to appear almost self-evident. If we conceive of a body moving in empty space, we can think of no reason why it should alter its path or its rate of motion in any way whatever.

(2) The second law presents, first, the proposition on which the measurement of force depends; and, secondly, states the identity of the direction of the change of motion with the direction of the force. Motion is here synonymous with momentum as before defined. The first proposition we have already employed in deriving the formula representing force. The second, with the

further statement that more than one force can act on a body at the same time, leads directly to a most important deduction respecting the combination of forces; for the parallelogram law for the resolution and composition of velocities being proved, and forces being proportional to and in the same direction as the velocities which they cause in any given body, it follows, if any number of forces acting simultaneously on a body be represented in direction and amount by lines, that their resultant can be found by the same parallelogram construction as that which serves to find the resultant velocity. This construction is called the *parallelogram of forces.*

In case the resultant of the forces acting on a body be zero, the body is said to be in *equilibrium.*

(3) When two bodies interact so as to produce, or tend to produce, motion, their mutual action is called a *stress.* If one body be conceived as acting, and the other as being acted on, the stress, regarded as tending to produce motion in the body acted on, is a force. The third law states that all interaction of bodies is of the nature of stress, and that the two forces constituting the stress are equal and oppositely directed.

27. Constrained Motion.—One of the most interesting applications of the third law is to the case of *constrained motion.* If the motion of a particle be restricted by the requirement that the particle shall move in a particular path, it is said to be constrained. If the velocity of the particle at a point in the path, at which the radius of curvature is r, be v, its acceleration toward the centre of curvature is $\frac{v^2}{r}$, and the force which must act on it in that direction is $\frac{mv^2}{r}$. However this force is applied, whether by a pull toward the centre or by a push or pressure from the body determining the path, or by the action of the forces which bind the particle to others moving near it, the reaction of the particle will in every case be equal to $\frac{mv^2}{r}$, and will be directed

outward along the normal to the path. This reaction is sometimes called a *centrifugal force*. There are certain cases in which it may be treated as if it were a real force, determining the motion of a body.

28. Work and Energy.—If the point of application of a force F move through a distance s, making the angle α with the direction of the force, the product $Fs \cos \alpha$ is defined as the *work* done by the force during the motion. If the force or the angle between the direction of the force and the displacement vary during the displacement, the work done may be found by dividing the path of the point into portions so small that $F \cos \alpha$ may be considered constant for each one of them. By forming the product $Fs \cos \alpha$ for each portion of the path, and adding all such products, the work done in the path is obtained.

In the defined sense of the term, no work is done upon a body by a force unless it is accompanied by a change of position, and the amount of work is independent of the time taken to perform it. Both of these statements need to be made, because of our natural tendency to confound work with conscious effort, and to estimate it by the effect on ourselves.

If work be done upon a particle which is perfectly free to move, its velocity will increase. In this case the force F is measured by ma, where m is the mass of the particle and a its acceleration. We may suppose that the particle has the velocity v_0 when it enters upon the distance s, and that the distance s coincides with the direction of the force. Using equation (10), we then have

$$Fs = mas = \tfrac{1}{2}mv^2 - \tfrac{1}{2}mv_0^2. \tag{20}$$

The product $\frac{1}{2}mv^2$ is called the *kinetic energy* of the particle. The equation shows that the work done upon the particle by a constant force is equal to the kinetic energy which it gains during the motion. If the direction of the motion or the magnitude of the force vary, we may divide the path into small portions, for each of which the force may be considered constant. Forming the equation just proved for each of these portions and adding the

equations thus obtained, we obtain for this general case the same result as that already obtained for the special case. The forces introduced by constraints need not be considered, since they are always perpendicular to the path, and so do no work.

When several forces act at a point, the work done by them during any small displacement of the point is equal to the work done by their resultant; for the sum of the projections of all the forces on the line of direction of the resultant is equal to the resultant, and the sum of the projections of each of these projections upon the direction of motion or the projection of the resultant upon the direction of motion is equal to the sum of the projections of each force upon the direction of motion. If, then, several forces do work on a particle, the kinetic energy gained by the particle will be equal to $Rs \cos \alpha$, where R is the resultant of the forces, and α the angle between its direction and the direction of the displacement s. Let us suppose that the forces are so related that $R = 0$. Then the work done by one of the forces must be equal and opposite to that done by the others, the particle will move with a constant velocity, and no kinetic energy will be gained. If any of the forces against which work is done are such that they depend only upon the position of the particle in the field, the work that is done against these forces is equal to that which is done by them if the particle traverse the path in the opposite direction. Such forces are called *conservative forces.* Other forces, which are not functions of the position of the particle only, but depend on its motion or some other property, are called *non-conservative forces.* When a particle acted on by conservative forces is so displaced that work is done against those forces, it is said to have acquired *potential energy.* The measure of the potential energy acquired is the work done against the conservative forces.

Energy is frequently defined as the capacity for doing work. The propriety of this definition is obvious in the case of potential energy; for the particle, acted on by conservative forces, and left free, will move under the action of these forces, and they will

thereby do work. The particle possessing kinetic energy has also the capacity for doing work, for, in order to bring it to rest, the amount of work given by the formula $Fs = \frac{1}{2}mv^2$ must be done upon it.

The unit of work and energy is the work done by a unit force upon a particle while it is displaced in the direction of the force through unit distance.

The dimensions of energy are ML^2T^{-2}, the same as those of work. Since the square of a length cannot involve direction, it follows that energy is a quantity independent of direction and is not a vector quantity.

The practical unit of work and energy is the *erg*.

It is the work done by a force of one dyne, in moving its point of application in the line of the force through a space of one centimetre:

Or, it is the energy of a body so conditioned that it can exert the force of one dyne through a space of one centimetre:

Or, it is the energy of a mass of two grams moving with unit velocity.

29. Bodies, Density.—The particle with which we have been dealing hitherto has no counterpart in Nature. In our experience we have to deal with extended *bodies* or systems of bodies, and the description of their motions and of the way in which forces act on them is more complicated than the corresponding description for the ideal particle. The notion of the particle is nevertheless of great utility: we may in the first place consider bodies as composed of numbers of these particles or as being systems of particles; and, in the second place, we may to some extent describe the motion of bodies by comparison with the motion of a particle.

It is, however, often convenient to be able to represent the mass of a body as distributed continuously throughout its volume. In that case we make use of a special concept, the *density*. To define it we suppose the particles of the body so distributed that each unit volume in the body contains the same number of them.

The density is then defined as the ratio of the mass of the body to its volume, or as the mass contained in a unit of volume. By supposing the mass of the body uniformly distributed throughout its volume, so that the ratio of mass to volume has the same value no matter how small the volume is, we may represent the mass contained in any infinitesimal volume by the product of the density and the volume. The concept of density used in this way is an artificial one, and the validity of the results obtained by it is due to the fact that the particles constituting a body are so small that their distribution is practically uniform in a homogeneous body in any volume which can be examined by experiment.

The formula for density is $D = \frac{M}{V}$, and the dimensions are $[D] = ML^{-3}$. The unit of density is the density of a homogeneous body so constituted that unit of mass is contained in unit of volume.

By using the hypothesis of a continuous distribution of matter in a body, we may define the density at a point in a body which is not homogeneous as the ratio of the mass contained in a sphere described about that point as centre to the volume of the sphere, when that volume is diminished indefinitely.

30. Centre of Mass.—The *centre of mass* of two particles is defined as the point which divides the straight line joining the particles into two segments, the lengths of which are inversely proportional to the masses of the particles at their extremities.

Thus if A and B be the positions of the two particles of which the masses are m_a and m_b respectively, then the point C, lying on the line joining A and B, is the centre of mass if it divide AB so that $m_a \cdot AC = m_b \cdot BC$.

The centre of mass of more than two particles is found by finding the centre of mass of two of them, supposing a mass equal to their sum placed at that centre, finding the centre of mass of this ideal particle and a third particle, and proceeding in a similar way until all the particles of the system have been brought into combination. The final centre thus found is the centre of mass of the

system. The point thus determined is independent of the order in which the particles are taken into combination; it is a unique point, and depends only on the positions of the particles and their masses.

The centre of mass may be defined analytically as follows: Let the particles m_1, m_2, ... be referred to a system of rectangular coordinates. The coordinates ξ, η, ζ of the centre of mass are then given by the equations

$$\left.\begin{aligned}\xi &= \frac{m_1x_1 + m_2x_2 + \ldots}{m_1 + m_2 + \ldots} = \frac{\Sigma mx}{\Sigma m};\\ \eta &= \frac{m_1y_1 + m_2y_2 + \ldots}{m_1 + m_2 + \ldots} = \frac{\Sigma my}{\Sigma m};\\ \zeta &= \frac{m_1z_1 + m_2z_2 + \ldots}{m_1 + m_2 + \ldots} = \frac{\Sigma mz}{\Sigma m}.\end{aligned}\right\} \qquad (21)$$

These equations are evidently consistent with the former definition of the centre of mass, if we remember that if the line joining any two particles be projected on one of the axes, the segments into which it is divided by the centre of mass of the two particles will be in the same ratio after projection as before. Consider the two particles m_1 and m_2, and denote the coordinate of their centre of mass by ξ. Then from the former definition of the centre of mass we have $m_1(\xi - x_1) = m_2(x_2 - \xi)$, from which $\xi = \frac{m_1x_1 + m_2x_2}{m_1 + m_2}$. This demonstration can easily be extended to include all the particles of the system.

If some of the particles of the system be in motion, the centre of mass will, in general, also move. Its velocity is determined by the velocities of the separate particles. Let ξ_0, η_0, ζ_0, represent the coordinates of the centre of mass at the time t_0, while ξ, η, ζ represent its coordinates at a later time t. The component velocities of the centre of mass are then given by the limit of the ratios $\frac{\xi - \xi_0}{t - t_0}$, $\frac{\eta - \eta_0}{t - t_0}$, $\frac{\zeta - \zeta_0}{t - t_0}$. Using the equations which define the coordinates of the centre of mass, we have:

$$\left.\begin{aligned}\frac{\xi - \xi_0}{t - t_0}\Sigma m &= \Sigma\frac{m(x - x_0)}{t - t_0};\\ \frac{\eta - \eta_0}{t - t_0}\Sigma m &= \Sigma\frac{m(y - y_0)}{t - t_0};\\ \frac{\zeta - \zeta_0}{t - t_0}\Sigma m &= \Sigma\frac{m(z - z_0)}{t - t_0}.\end{aligned}\right\} \qquad (22)$$

The terms on the right are the components of momentum of the separate particles, and the equations express the law that the velocity of the centre of mass of a system of particles is equal to the resultant obtained by compounding the momenta of the separate particles and dividing it by the sum of all the masses of the system.

Representing the component velocities of the centre of mass by U, V, W, and those of the separate particles by u, v, w, the rule just given may be expressed by $U\Sigma m = \Sigma mu$, $V\Sigma m = \Sigma mv$, $W\Sigma m = \Sigma mw$.

If the velocities of some or all of the particles vary, the velocity of the centre of mass will in general vary also. Its acceleration depends upon the accelerations of the separate particles. Letting U and U_0, etc., represent the component velocities at the times t and t_0, we may express the component accelerations of the centre of mass by

$$\left.\begin{aligned}\frac{U - U_0}{t - t_0}\Sigma m &= \Sigma\frac{m(u - u_0)}{t - t_0};\\ \frac{V - V_0}{t - t_0}\Sigma m &= \Sigma\frac{m(v - v_0)}{t - t_0};\\ \frac{W - W_0}{t - t_0}\Sigma m &= \Sigma\frac{m(w - w_0)}{t - t_0}\end{aligned}\right\} \qquad (23)$$

The terms on the right represent the components of the forces which act on each particle of the system, and the equations express the law that the acceleration of the centre of mass of a system of particles is equal to the resultant of all the forces which act on the separate particles divided dy the sum of the masses of the particles. This law may be otherwise expressed by saying that the acceleration of the centre of mass is the same as that which would be given to

a particle having a mass equal to the sum of all the masses if it were acted on by a force equal to the resultant of all the forces.

Forces which act between particles belonging to the same system are called *internal forces*; such forces do not affect the motion of the centre of mass, for, by Newton's third law of motion, they always occur in pairs, of which the two members are equal and opposite. They therefore contribute nothing to the resultant force, and so do not influence the acceleration of the centre of mass. If the only forces which act be internal forces, the acceleration of the centre of mass is zero and the momentum of the system remains constant. This principle is known as the *conservation of momentum.*

31. Kinetic Energy of a System of Particles.—The kinetic energy of a system of particles may also be expressed in terms of the velocity of the centre of mass. Represent by u, v, w the components of velocity of each particle, by U, V, W the components of velocity of the centre of mass, and by a, b, c the components of velocity of each particle relative to the centre of mass. We have then

$$\begin{aligned} u_1 &= U + a_1, & u_2 &= U + a_2, \ldots \\ v_1 &= V + b_1, & v_2 &= V + b_2, \ldots \\ w_1 &= W + c_1, & w_2 &= W + c_2, \ldots \end{aligned}$$

The kinetic energy of the particle m_1 is $\frac{1}{2}m_1(u_1^2 + v_1^2 + w_1^2)$, and the kinetic energy of all the particles or of the system is the sum of the similar expressions obtained for each particle of the system. Substitute in the equation for the kinetic energy the values of u^2, v^2, w^2. We consider first the values of u^2. We have

$$u_1^2 = U^2 + a_1^2 + 2a_1U, \quad u_2^2 = U^2 + a_2^2 + 2a_2U, \quad \ldots$$

Multiplying by $\frac{1}{2}m$ and adding, we obtain

$$\begin{aligned} \Sigma\tfrac{1}{2}mu^2 = \tfrac{1}{2}U^2(m_1 + m_2 + \ldots) + \tfrac{1}{2}m_1a_1^2 + \tfrac{1}{2}m_2a_2^2 + \ldots \\ + U(m_1a_1 + m_2a_2 + \ldots). \end{aligned}$$

Now since a_1, a_2, . . . are referred to the centre of mass as origin, and since in that case the coordinates of the centre of mass are zero, the sum $m_1a_1 + m_2a_2$ . . . must equal zero. If the expres-

sions for $\frac{1}{2}m_1v_1^2 + \frac{1}{2}m_2v_2^2 \ldots, \frac{1}{2}m_1w_1^2 + \frac{1}{2}m_2w_2^2 \ldots$ be formed in a similar manner, and added to the expression just obtained, we have on the left the sum of the kinetic energies of the particles, and on the right the expression

$$\tfrac{1}{2}(U^2+V^2+W^2)\Sigma m+\tfrac{1}{2}m_1(a_1^2+b_1^2+c_1^2)+\tfrac{1}{2}m_2(a_2^2+b_2^2+c_2^2)+\ldots$$

The first of these terms expresses the kinetic energy of a mass equal to the sum of all the masses moving with the velocity of the centre of mass. The other terms express the kinetic energies of the separate particles moving with their velocities relative to the centre of mass. We therefore arrive at the following rule: The kinetic energy of a system of particles is equal to the kinetic energy of a mass equal to the sum of all the masses moving with the velocity of the centre of mass, plus the kinetic energies of the separate masses moving with their velocities relative to the centre of mass.

32. Work done by Forces on a System of Particles. Potential Energy.—The forces which act on the particles of a system may be classified as *external* and *internal forces*. The external forces arise from the action of bodies outside the system, the internal forces from action between parts of the system. If the resultant of all the forces which act on any one particle be considered as the force which acts on that particle, the particle will acquire kinetic energy, given by the formula $Fs = \frac{1}{2}mv^2 - \frac{1}{2}mv_0^2$, already established (§ 28). If, however, we consider the resultant of the external forces acting on the particle as producing kinetic energy and doing work against the internal forces which act on the particle, the work done by the former will be equal to the kinetic energy gained by the particle plus the work done against the latter. If the internal forces be conservative, the work done against them can be recovered when the external forces cease to act. The action of the external forces in that case gives to each particle potential energy. In case the external forces equilibrate the internal forces for each particle, the velocities of the particles remain constant, no kinetic energy is gained, and the energy given to the system by the work done is wholly potential. In any case the energy gained

by the system is equal to the work done on it by the external forces. If no external forces act on a system, its energy remains constant, however the velocities of the separate particles may change in consequence of the action of internal forces.

A *rigid body* is one in which the particles retain the same relative positions. Whatever internal forces act between the particles, they are equilibrated by others due to the reactions in the system. The internal forces can therefore do no work, and the internal energy of such a body is wholly kinetic energy.

33. Conservation of Energy.—The theorem stated in the last section is the simplest illustration of the general principle known as the conservation of energy. If no external forces act on a system, and if the internal forces be conservative, the sum of the kinetic and potential energies of the system remains constant. In many operations in Nature, however, the internal forces are not all conservative, and the theorem just stated no longer holds true. Experiment has shown that when non-conservative forces act, other forms of energy are developed, which cannot as yet be expressed as the potential and kinetic energies of masses, and that if these forms of energy be taken into account, the sum of all the energies of the system remains constant so long as no external forces act on it. This principle is called the principle of the *conservation of energy*. It may be used as a working principle in solving questions in mechanics, and finds a very wide application in all departments of physical science. The evidence for it will appear in connection with many of the topics which are subsequently treated.

34. Systems to be Studied.—The description of the motions of a system of particles which are free to move among themselves, and between which forces act, cannot in most cases be given. Certain general theorems relating to this general case can be found, but the conditions which determine the individual motions of the particles are so complicated that they cannot be brought into a form suitable for mathematical discussion, and hence the motion of the system cannot be completely described. There are two cases, however, of very general character, in which, by the aid of

certain limitations assumed for the system, we are able fully to describe its motions. The first of these is that of a pair of bodies which act on each other with a force, the direction of which is in the line joining the bodies. This case, known as the problem of two bodies, may be completely solved. The problem of three bodies can be solved only approximately, under certain limitations as to the relative magnitudes of the bodies. The second case is that in which the system forms a rigid body. While no truly rigid bodies exist in Nature, yet the changes of shape which most solids undergo under the action of ordinary forces are so slight in comparison with their dimensions that in many cases we may consider such solids as rigid, and illustrate the theorems relating to rigid bodies by experiments made upon solids. We shall first examine the motion of rigid bodies, and we shall limit ourselves to the case in which the motions of any one particle of the body always take place in one plane. By thus restricting the problem, it is possible to obtain the most essential facts connected with the motions of rigid bodies without the use of advanced mathematical methods.

35. Impact.—The changes in motion impressed upon bodies by their impact with others depend upon so many conditions that they present complications which render the discussion of them impossible in this book. We will consider, however, the simple case of the impact of two spheres, the centres of which are moving in the same straight line. We call the masses of the two spheres m_1 and m_2 and their respective velocities u_1 and u_2. The two spheres constitute a system for which the velocity of the centre of mass is given by

$$(m_1 + m_2) V = m_1 u_1 + m_2 u_2. \tag{25}$$

The bodies on impact are momentarily distorted, and a force arises between them tending to separate them, the magnitude of which depends upon the elasticity of the bodies. The velocity of the centre of mass will remain uniform, whatever be the forces acting between the bodies, and the momenta of the two bodies relative to the centre of mass, both before and after impact, will be equal and opposite. Call the velocities of the bodies after impact v_1 and

v_2. We then nave $m_1(u_1 - V) = m_2(V - u_2)$ and $m_1(V - v_1) = m_2(v_2 - V)$. That these equations may both be true we must have $\frac{V - v_1}{u_1 - V} = \frac{v_2 - V}{V - u_2} = e$, an experimental constant, called the *coefficient of restitution*. The coefficient e depends upon the elasticity of the bodies and their mode of impact. It has been shown by experiment to be always less than unity. From these equations we deduce

$$e(u_1 - u_2) = v_2 - v_1. \tag{26}$$

Combining this equation with the equation for the velocity of the centre of mass, we obtain for the velocities v_1 and v_2 after impact the equations

$$\left.\begin{aligned} v_1 &= V - \frac{m_2}{m_1 + m_2} e(u_1 - u_2); \\ v_2 &= V + \frac{m_1}{m_1 + m_2} e(u_1 - u_2). \end{aligned}\right\} \tag{27}$$

The kinetic energy before impact equals $\frac{1}{2}m_2u_2^2 + \frac{1}{2}m_2u_2^2$. The kinetic energy after impact equals $\frac{1}{2}m_1v_1^2 + \frac{1}{2}m_2v_2^2$. Substituting in this last expression the values just obtained for v_1 and v_2 and reducing, we obtain for the kinetic energy after impact

$$\frac{(m_1 + m_2)(1 - e^2)V^2}{2} + \frac{e^2(m_1u_1^2 + m_2u_2^2)}{2}.$$

By subtracting this from the kinetic energy before impact we find that the loss of kinetic energy by impact is

$$\frac{m_1m_2}{m_1 + m_2} \cdot \frac{1 - e^2}{2}(u_1 - u_2)^2. \tag{28}$$

If the bodies are such that $e = 0$, or such that the velocities after impact are both equal to the velocity of the centre of mass, they are called *inelastic* bodies; the kinetic energy lost by their collision is $\frac{m_1m_2}{m_1 + m_2} \cdot \frac{(u_1 - u_2)^2}{2}$. If, on the other hand, $e = 1$, so that the velocities after impact relative to the centre of mass are equal to those before impact but of opposite sign, the bodies are called *perfectly elastic* bodies. In this case no kinetic energy is lost by

the collision. These extreme values of e are never exhibited by real bodies, though the value $e = 0$ may be closely approached in many instances. No body has a value of e that is even appreciably equal to 1, so that there is always a loss of kinetic energy by impact. The energy thus lost is transformed into other forms of energy, principally into heat.

36. Displacement of a Rigid Body.—Under the limitation that we have set, that the points of the body shall move only in parallel planes, it is manifest that the motion of the body is completely given if the motion of its section by any one plane be given. In describing the displacement of a body under these limitations we need only describe the displacement of one of its sections by one of the planes in which the motion occurs. It is furthermore clear that the motion of this section will be completely described if the motion of any two points in it or of the line joining them be given.

When a body is so displaced that each point in it moves in a straight line through the same distance, its displacement is called a *translation*. When the points of the body describe arcs of circles which have a common centre, its displacement is called a *rotation*. Any displacement of a body may be effected by a translation combined with a rotation. To show this, let AB (Fig. 11) represent the initial position of a line in the body, $A'B'$ its final position. The transfer from the initial to the final position may be effected by a translation of the line AB to such a position that the point C, which may be any point in the body, coincides with the corresponding point C'. Taking this point C' as the centre, a rotation through an angle θ, which is the same whatever point be chosen for C, will bring the line into its final position. While the angle of rotation is the same whatever point be chosen for C, the translation which brings C into coincidence with C' will differ for different positions of C.

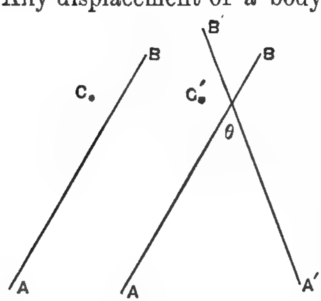

FIG. 11.

If the line AB be rotated through the angle θ about any point in it, and if then another point in it be taken and the line rotated about that point through an angle $-\theta$, the result is a translation of the line AB. We may therefore substitute for a rotation about one point a translation and an equal and opposite rotation about another properly chosen point.

By the following construction it is always possible to find a point O in the plane in which AB moves, such that a pure rotation of AB about it will bring the body from its initial to its final position.

Join AA', BB' (Fig. 12), and bisect the lines AA' and BB' at the points C and D. At those points erect perpendiculars which will intersect at the point O. Join OA, OB, OA', OB'. By the geometry of the figure the triangles AOB and $A'OB'$ are similar, and adding to their equal angles at O the common angle $A'OB$, we have $AOA' = BOB'$. Hence a rotation through the angle $AOA' = BOB'$ will transfer AB to $A'B'$. The perpendicular through O may be called the *axis* of rotation. This construction fails when the initial and final positions of AB are parallel.

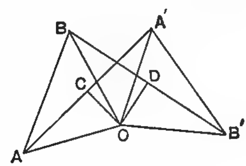

FIG. 12.

37. Kinetic Energy of a Rotating Body.—Let r represent the distance of any particle of the body, of mass m, from the axis about which the body rotates, and ω its angular velocity about that axis. Then the kinetic energy of this particle is $\frac{1}{2}mr^2\omega^2$, and the kinetic energy of the rotating body is $\frac{1}{2}\omega^2\Sigma mr^2$. In § 36 we have shown that we may replace a rotation by a translation and a rotation of the same amount about another axis. Since velocities are measured by the displacements of the moving particle which occur in the same interval of time, it is also possible to replace an angular velocity by a velocity of translation and an equal angular velocity in the opposite sense about another axis. We choose for the new axis that passing through the centre of mass, at the distance R from the original axis. The velocity of the centre of mass is then $R\omega$. We represent by l the distance of the mass m from the axis passing

through the centre of mass. The kinetic energy of the body rotating about this centre is $\frac{\omega^2}{2}\Sigma ml^2$, and the kinetic energy of the whole body moving with the velocity of the centre of mass is $\frac{1}{2}\omega^2 R^2 \Sigma m$. By § 31 we have

$$\tfrac{1}{2}\omega^2 \Sigma mr^2 = \tfrac{1}{2}\omega^2 R^2 \Sigma m + \tfrac{1}{2}\omega^2 \Sigma ml^2. \tag{29}$$

When a rigid body is so small that its kinetic energy due to its rotation about its centre of mass is negligible in comparison with that due to its translation, it is called a *particle*. This definition supplements that of § 25.

38. Moment of Inertia.—The expression Σmr^2 is called the *moment of inertia* of the body about the axis from which r is measured. The formula just obtained shows that the moment of inertia about any axis is equal to the moment of inertia about a parallel axis passing through the centre of mass plus the moment of inertia of a particle of which the mass is equal to the mass of the body placed at the centre of mass.

The moment of inertia depends entirely upon the magnitude of the masses making up the body and their respective distances from the axis. If the mass of the body be distributed so that each element of volume contains a mass proportional to the volume of the element, the moment of inertia then becomes a purely geometrical magnitude, and may be found by integration.

It is evident that it is always possible to find a length k such that $k^2\Sigma m = \Sigma mr^2$. This length k is called the *radius of gyration* of the body about its axis.

The moment of inertia of any body, however irregular in form or density, may be found experimentally by the aid of another body of which the moment of inertia can be computed from its dimensions. We will anticipate the law of the pendulum—which has not been proved—for the sake of clearness. The body of which the moment of inertia is desired is set oscillating about an axis under the action of a constant force. Its time of oscillation is, then, $t = \pi\sqrt{\frac{I}{f}}$,

where I is the moment of inertia and f a constant depending on the magnitude of the force.

If, now, another body, of which the moment of inertia can be calculated, be joined with the first, the time of oscillation changes to $t' = \pi\sqrt{\dfrac{I+I'}{f}}$, where I' is the moment of inertia of the body added. Combining the two equations, we obtain, as the value of the moment of inertia desired,

$$I = \frac{I't^2}{t'^2 - t^2}. \tag{30}$$

39. Rotation about a Fixed Point.—Suppose a body so conditioned that its only motion is a rotation about the fixed point O (Fig. 13). Suppose the force F applied at a point in the body, which moves under the action of the force through the infinitesimal distance QR. This motion is a rotation about the point O through the angle $\phi = \dfrac{QR}{OQ}$. The work done by the force during this rotation is

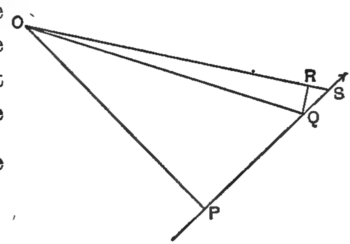

FIG. 13.

$$W = F.\frac{QR}{QS}\cdot QR.$$

Since, in the limit, when QR and QS are infinitesimal, the triangles QRS and OPQ are similar, $\dfrac{QR}{QS} = \dfrac{OP}{OQ}$, and hence

$$W = F\,.\,OP\,.\,\frac{QR}{OQ} = F\,.\,OP\,.\,\phi.$$

The work thus done is equal to the kinetic energy gained by the rotating body, or to $\frac{1}{2}\omega^2 I$, where I is the moment of inertia of the body and ω the angular velocity which it gains during the motion. Now $\omega^2 = 2\alpha\phi$, where α is the angular acceleration (§ 20), and hence

$$F\,.\,OP = I\alpha. \tag{31}$$

The product $F\,.\,OP$, or the product of the force and the perpendicular let fall from the axis of rotation upon the line of direction

of the force, is called the *moment* of the force about that axis. Since, if several forces act on the body, they each contribute their share to the angular acceleration produced, we have also $\Sigma F \,.\, OP = I\alpha$.

40. Principle of Moments.—If the angular velocity of the body be constant, we have $\alpha = 0$, and hence $\Sigma F \,.\, OP = 0$. The body is then said to be in equilibrium about the fixed axis. Hence a body free to rotate about a fixed axis is in equilibrium if the sum of the moments of the forces which tend to turn it in one sense is equal to the sum of the moments of the forces which tend to turn it in the opposite sense. This theorem is called the *principle of moments.*

41. Principle of Work.—If the body rotate uniformly about a fixed axis, it does not gain angular velocity, and we have $\omega = 0$, and therefore $\Sigma F \,.\, OP \,.\, \phi = 0$. Now $OP \,.\, \phi$ is the distance traversed by the point of application of the force, and this is proportional to the velocity v with which that point of application moves. Therefore $\Sigma Fs = 0$, or $\Sigma Fv = 0$. The body is in equilibrium about a fixed axis when the positive work done upon it by some of the forces applied to it during any small displacement is equal to the negative work done by the other forces upon it. The expression $Fv = \frac{Fs}{t}$ measures the rate at which work is done by the force, and the condition of equilibrium may be otherwise stated by saying that the rotating body is in equilibrium when the rate at which positive work is done upon the body equals the rate at which negative work is done upon it.

42. Couples.—A combination of two forces which are equal and oppositely directed, but not in the same straight line, is called a *couple.* The sum of their moments (Fig. 14) is $F \,.\, OQ - F \,.\, OP = F \,.\, PQ$, and is manifestly the same wherever the forces are applied in the body, provided the distance PQ remains the same. PQ is called the arm of the couple.

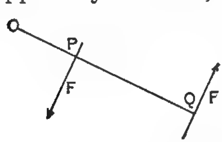

FIG. 14.

Since the effect of different forces in producing rotation is the same if the sum of their moments is the same, it is also clear that the

couple may be replaced by any other couple if the product $F \cdot PQ$ is the same for both. The couple will have the same moment whatever point be chosen as the fixed point. Since the two forces which constitute the couple are equal and opposite, their resultant is zero, and therefore no single force can be found which will equilibrate a couple.

43. Movement of a Free Body.—Whatever forces act on a free body, they may always be reduced to a single force applied at the centre of mass and to a single couple which will produce rotation about the centre of mass. Let F (Fig. 15) be any one of these forces. Apply the equal and opposite forces F and $-F$ at the centre of mass O. These two forces, having no resultant, will not affect the motion of the body. The force F applied at the centre of mass will determine the effect of the original force F in producing translation of the centre of mass. The original force F and the force $-F$ constitute a couple, the arm of which is OP and which produces rotation about the centre of mass. Treating all other forces applied to the body in a similar way, we have finally an assemblage of forces applied at the centre of mass, the resultant of which determines the acceleration of the centre of mass or the translation of the body, and a set of couples, equivalent to a single couple of which the moment is equal to the sum of their moments, which produces rotation about the centre of mass.

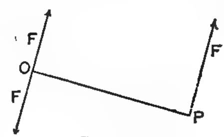

FIG. 15.

A free body is in equilibrium, or undergoes no change in the velocity of its centre of mass or in its rotation, when the resultant R of the forces applied to it vanishes, and the moment of couple to which the moments of these forces about the centre of mass may be reduced also vanishes, that is, the body is in equilibrium when $R = 0$ and $\Sigma F \cdot OP = 0$.

44. Centre of Percussion.—We will illustrate the foregoing principles by considering the motion of a free body to which a force is applied for a very short time, during which it may be considered constant. The force is supposed to act in the plane containing the centre of mass of the body. Then, as has just been shown, the ac-

celeration of the centre of mass is given by $F = Ma$, and the angular acceleration about the centre of mass by $F . R = I\alpha$, where R is the distance from the centre of mass to the line of the force. The actual acceleration of any point of the body about the centre of mass, due to this angular acceleration, is αx, where x is the distance of the point from the centre of mass. The total acceleration of any point on the line of R is $a \pm \alpha x = \frac{F}{M} \pm \frac{F . R}{I} x$, the positive or negative sign being used according as the point lies in R itself, or in its prolongation through the centre of mass. If $a - \alpha x = 0$, the point considered is at rest. The condition that the point is at rest is therefore $\frac{1}{M} - \frac{Rx}{I} = 0$, or

$$x = \frac{I}{MR}. \tag{32}$$

The movement of the body will, therefore, not be altered if a fixed axis be passed through this point. If the body be considered as free to rotate about this axis, the point of application of the force, which is distant $R + x$ from the axis, and which is such that the force there applied will occasion no stress on the axis, is called the *centre of percussion*. We have

$$x + R = \frac{I}{MR} + R = \frac{I + MR^2}{MR}. \tag{33}$$

By § 38, $I + MR^2$ is the moment of inertia of the body about the axis of rotation, so that the distance from the axis of rotation to the centre of percussion is equal to the moment of inertia of the body divided by MR. The product MR is sometimes called the *static moment*.

45. Mechanical Powers.—There are certain simple cases of the combination of forces in accordance with the foregoing principles which are of especial importance because of their application in the construction of machines. They are generally called *the mechanical powers*.

They are all designed to enable us, by the application of a certain force at one point, to obtain at another point a force, in general

not equal to the one applied. Six mechanical powers are usually enumerated—the lever, pulley, wheel and axle, inclined plane, wedge, and screw.

(1) *The Lever* is any rigid bar, of which the weight may be neglected, resting on a fixed point called a *fulcrum*. From the proposition in § 40 it may be seen that if forces be applied to the ends of the lever there will be equilibrium when the resultant passes through the fulcrum, and the moments of force about the fulcrum are equal. Hence, if the forces act in parallel lines, it follows that the force at one end is to the force at the other end in the inverse ratio of the lengths of their respective lever-arms. If l and l' represent the lengths of the arms of the lever, and P and P' the forces applied to their respective extremities, then $Pl = P'l'$.

The principle of the equality of action and reaction enables us to substitute for the fulcrum a force equal to the resultant of the two forces. We have then a combination of forces as represented in the diagram (Fig. 16). Plainly, any one of these forces may be considered as taking the place of the fulcrum, and either of the others the *power* or the *weight*.

FIG. 16.

The lever is said to be of the first kind if R is fulcrum and P power, of the second kind if P' is fulcrum and P power, of the third kind if P is fulcrum and R power.

(2) *The Pulley* is a frictionless wheel, in the groove of which runs a perfectly flexible, inextensible cord.

If the wheel be on a fixed axis, the pulley merely changes the direction of the force applied at one end of the cord. If the wheel be movable and one end of the cord fixed, and a force be applied to the other end parallel to the direction of the first part of the cord, the force acting on the pulley is double the force applied: for the stress on the cord gives rise to a force in each branch of it equal to the applied force; each of these forces acts on the wheel, and since the radii of the wheel are equal, the resultant of these two forces is a force equal to their sum applied at the centre of the

wheel. From these facts the relation of the applied force to the force obtained in any combination of pulleys is evident.

(3) *The Inclined Plane* is any frictionless surface, making an angle with the line of direction of the force applied at a point upon it. Resolving the force P (Fig. 17), making an angle ϕ with the normal to the plane, into its components $P \cos \phi$ and $P \sin \phi$ perpendicular to and parallel with the plane, $P \sin \phi$ is alone effective to produce motion. Consequently, a force $P \sin \phi$ acting parallel to the surface will balance a force P, making an angle ϕ with the normal to the surface. If the plane be taken as the hypothenuse of a right-angled triangle ABC, of which the base AB is perpendicular to the line of direction of the force, then, by similarity of triangles, the angle BAC equals ϕ: whence the force obtained parallel to AC is equal to the force applied multiplied by the sine of the angle of inclination of the plane. If the components of the force applied be taken, the one, as before, perpendicular to the plane AC, and the other parallel to the base AB, the force obtained parallel to AB is equal to the force applied multiplied by the tangent of the angle of inclination of the plane.

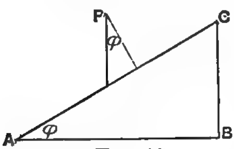

Fig. 17.

(4) *The Wheel and Axle* is essentially a continuously acting lever.

(5) *The Wedge* is made up of two similar inclined planes set together, base to base.

(6) *The Screw* is a combination of the lever and the inclined plane.

The special formulas expressing the relations of the force applied to the force obtained by the use of these combinations are deduced from those for the more elementary mechanical powers.

Any arrangement of the mechanical powers, designed to do work, is called a *machine*. The more nearly the value of the work done approaches that of the energy expended, the more closely the machine approaches perfection. The elasticity of the materials we are compelled to employ, friction, and other causes which mod-

ify the conditions required by theory, make the attainment of such perfection impossible.

The ratio of the useful work done to the energy expended is called the *efficiency* of the machine. Since in every actual machine there is a loss of energy in the transmission, the efficiency is always a proper fraction.

46. Application of the Principle of Work to the Mechanical Powers.

(1) *The Lever.*—If the lever be displaced through a small angle ϕ about its fixed point or fulcrum, the distances traversed by the points of application of the two forces which act on its ends, these forces being supposed parallel, are equal to $l\phi$ and $l'\phi$. The lever is in equilibrium, by the principle of work, when the work done by the force P during this rotation is equal and of opposite sign to that done by the force P', or when $Pl - P'l' = 0$. Hence equilibrium obtains, as already shown, when $Pl = P'l'$.

(2) *The Pulley.*—In any combination of pulleys the forces which are equivalent to the reaction of the parts of the system always occur in pairs, and the work done by the two members of each of these pairs is equal in magnitude but opposite in sign, so that all such forces may be neglected in the enumeration of those forces which are concerned in doing work in the system, and which must be considered in applying the principle of work. The only forces which need be considered are the external force W, the weight, and the external force P, the power, which is applied to equilibrate the weight. If w and p represent the distances passed over by W and P respectively, when the cord is drawn through the pulleys, equilibrium will obtain when $Ww = Pp$. The dependence of the ratio $\frac{W}{P}$ upon the number of pulleys and their arrangement may always be deduced from this equation.

(3) *The Inclined Plane.*—If the weight W which rests upon the plane be acted upon by the force P parallel with the surface of the plane, equilibrium will obtain when the work done by each of these forces, during any small displacement of the weight up or down

the plane, is the same. If ϕ be the angle between the direction of the force W and the normal to the surface of the plane, and s the distance traversed by the weight, this condition is fulfilled when $Ps - Ws \sin \phi = 0$. Hence the condition of equilibrium is $P = W \sin \phi$.

The principle of work is of special value in all such cases as those illustrated by the combination of pulleys, in which, however complicated the arrangement of the parts of the system which transfer the action of the applied force or power to the point of application of the other force, we know that the forces equivalent to the reactions between those parts occur in pairs, of which the members are equal and opposite, so that the work done during any displacement of the system by each of these pairs is zero; for in any such case equilibrium obtains when the work done by the one force equals the work done by the other.

47. Motion of a Rigid Body in Three Dimensions.—The motion of a rigid body which is not under the restriction hitherto imposed upon it, but which is free to move in all directions, is in many respects analogous to the motion already studied, though the details are necessarily more complicated. We will attempt no demonstration of the laws of the motion of a rigid body in the general case, but will limit ourselves to a short description of them.

Any displacement of a rigid body may always be replaced by a translation and a rotation about some axis; this may readily be seen by considering any simple example. By the use of an example, it will also appear that the direction of the axis does not in general coincide with the direction of the translation; it is, however, always possible to find a direction such that translation in that direction and rotation about an axis in that direction will produce the displacement required. An infinitesimal displacement may be produced, therefore, by an infinitesimal translation and a rotation about an axis in the direction of the translation, that is, by a motion resembling that of a screw when driven forward. The axis of rotation in this case, which will in general change its direction and position in space as the body traverses its path, is called the *instantaneous*

axis, and any infinitesimal motion of a body is a screw motion around the instantaneous axis.

The rotation of a rigid body free to move and acted on by no forces will be about an axis passing through the centre of mass. The kinetic energy of the rotating body depends on its moment of inertia about its axis of rotation, and this may be shown to depend upon the moments of inertia of the body about three axes in it at right angles to each other; these axes, called the *principal axes of inertia,* are such that the moment of inertia about one of them is the greatest, and that about another the least, that the body can have. If a body be set rotating around either of these two axes, it will continue to rotate about that axis forever, and its condition is stable; that is, if an infinitesimal change be made in the direction of its axis of rotation, this deviation will never become large. If it be set rotating about the third or mean axis, it will continue to rotate about that axis forever, but its condition is unstable; that is, if an infinitesimal change be made in the direction of the axis of rotation, this deviation will tend continually to increase, and will become finite. If the body be set rotating about any axis which is not coincident with one of the principal axes, the direction of the axis of rotation in the body changes continually. In the case of real bodies set in rotation and acted on by friction and other such forces, the tendency is for the body to rotate with increasing exactness around the axis of greatest moment of inertia.

In the study of the angular velocity of a rotating body we represent the axis of rotation by a line and the amount of the angular velocity by a length measured on that line. If we conceive of two angular velocities about intersecting axes, it may be shown that they are equivalent to a single angular velocity about another axis passing through the point of intersection; the amount of this angular velocity and the direction of the axis are determined by the parallelogram law. Manifestly this law may be applied to the composition and resolution of any number of angular velocities about axes which intersect at one point.

The resolution of angular velocities into their components is

well illustrated by the instrument called the *Foucault pendulum.* This pendulum is a heavy spherical bob suspended by a long cylindrical wire, so clamped that it is perfectly free to swing in any plane. If such a pendulum were set up at the equator and started swinging in a north and south line, it would be carried around by the earth in its rotation and would evidently continue to swing in the same line on the earth's surface. If, on the other hand, it were set up at the pole and set swinging along any line traced on the earth's surface, the oscillations of the pendulum would persist in the same plane in space, and the earth would turn around under it, so that the oscillations of the pendulum would immediately begin to deviate from the line on the earth along which the first oscillation took place; and if the oscillations were continued the extremities of the paths of the pendulum would describe the arc of a circle, and at the end of a day the pendulum would again swing in the line in which it started. If now the pendulum be swung at some intermediate point on the earth's surface the angular velocity of the earth can be resolved into two component angular velocities—one about the axis *OA* (Fig. 18), which coincides with the length of the pendulum when it is at rest, and the other about an axis *OB* at right angles to this. The angular velocity about the latter axis will have no effect on the line traced out by the swinging pendulum; the angular velocity about the former axis will occasion an apparent angular displacement. By the parallelogram law this angular velocity equals $\omega \sin \phi$, where ω is the angular velocity of the earth about its axis, and ϕ is the latitude. By experiments with such a pendulum this formula is verified and the rotation of the earth established by direct experiment, and by assuming the validity of this formula and determining the angular displacement of the pendulum an approximate value of the length of the day has been obtained.

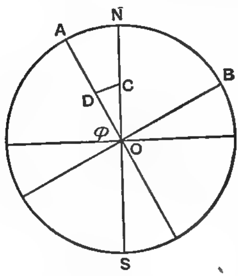

FIG. 18.

If a body in rotation about its axis of greatest moment of inertia be acted on by a couple tending to change the direction of rotation, that is, to give it an angular acceleration about another axis at right angles to this, the various reactions which arise from the changing directions of the parts of the body which are in motion occasion other couples which oppose the action of the applied couple. That is, the rotating body possesses a certain stability due to its rotation. This effect is illustrated by the instrument called the *gyroscope* and by the common top.

The construction of the gyroscope can best be understood by the help of the diagram (Fig. 19). The outermost ring rests in a

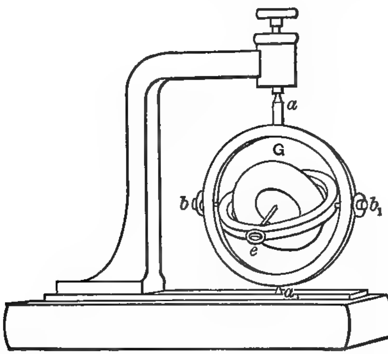

FIG. 19.

frame, and turns on the points a, a_1. The inner rests in the outer one, and turns on the pivots b, b_1 at right angles to the line of a, a_1. Within this ring is mounted the wheel G, the axle of which is at right angles to the line bb_1, and in a plane passing through aa_1. At the point e is fixed a hook, from which weights may be hung. It is evident that if the wheel be mounted on the middle of the axle the equilibrium of the apparatus is neutral in any position, and that a weight hung on the hook e will bring the axle of the wheel vertical, without moving the outer ring. If, however, the wheel be set in rapid rotation, with its axle horizontal, and a weight be hung on the hook, the whole system will revolve with a constant angular

velocity about the points a, a_1, and the axle of the wheel will remain horizontal.

Let Fig. 20 represent the rotating wheel of the former diagram, the axis being supposed to be nearly horizontal. If a weight be hung at the point e, it tends to turn the wheel about a horizontal axis CD. The direction of motion of the particles at A and B is not changed by this rotation, but the particles at C and D, and to a less extent all the other particles on the rim of the wheel, are forced to change their directions of motion. Now it has been shown (§ 27) that the change in the direction of motion of a particle is equivalent to a force $\frac{mv^2}{r}$, where m is the mass of the particle, v its velocity, and r the radius of the circle in which it moves. The reaction of the particle is directed outward along the normal to the curve; in the case of the particles considered at C and D, this reaction is directed to the right at C and to the left at D. These two forces, therefore, and all others like them due to the reactions of the other particles, combine to form a couple which tends to rotate the wheel about the axis AB. This rotation about AB gives rise to similar reactions at A and B, the reaction at A being directed to the left and at B to the right. These forces, and all other similar ones arising from the other particles of the wheel, combine to form a couple which tends to rotate the wheel about the axis CD in the opposite sense to that in which it is rotated by the weight at e. Thus the weight applied at e will produce a rotation about the vertical axis AB.

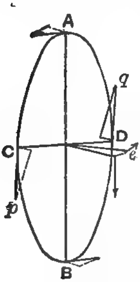

FIG. 20.

48. Central Forces.—We now turn to the consideration of the motion of a particle acted upon by a force always directed toward a fixed centre or a *central force*. Its motion will exhibit one peculiarity which is independent of the law of the central force. The radius drawn from the centre to the particle will always sweep out equal areas in equal times, whatever be the law of the force.

It is at once obvious that the motion of a particle, acted on by

any central force, will always lie in one plane—that containing its original direction of motion and that of the force.

Suppose that the moving particle, which starts from the point A (Fig. 21), and which in the time t moves over the distance AB, is acted on at the point B by an impulse directed toward the centre O, such that the particle is displaced toward O in the next equal time-interval t by the distance BE. If the particle were not acted on by the impulse at B it would continue to move in the line AB, and at the end of the second time-interval t it would reach the point D, the line BD being equal to the line AB. Join OD, draw DC parallel to OB, and EC parallel to BD; connect BC. Now the line BC is the resultant of the displacement BD, which the particle would have in consequence of its original motion, and the displacement BE, given to it by the impulse at B. It is therefore the path traversed by the particle in the second time-interval. But the triangles OCB and ODB, being on the same base and between the same parallels, are equal; and the triangles OBD and OAB, being on equal bases and having the same vertex O, are equal. Therefore the triangle OCB and the triangle OAB, described in equal time-intervals, are equal. If now the intervals into which the whole time is divided become infinitely small, in the limit the broken line ABC approaches indefinitely near to a curve, and the areas swept out in equal times by the radius vector drawn to the curve are equal.

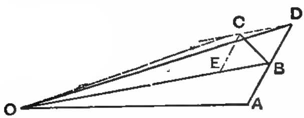

FIG. 21.

If a line be drawn tangent to the path at any point, and a perpendicular, p, drawn to it from the centre, the area swept out by the radius vector as the particle describes a small distance $s = vt$, where v is its velocity and t the time in which s is described, is given by $\frac{1}{2}vpt$, and since the areas described in equal times are equal, the product vp is constant for all parts of the path.

49. Central Force Proportional to the Radius Vector.—If the central force which acts on the particle be proportional to the distance of the particle from the centre, the path which it describes

is in general an ellipse, with its centre at the centre of force. For, the force acting along the radius vector may be resolved into two components along the two axes, which will be proportional to the displacement of the particle from the axes. Each of these components will cause proportional accelerations along the axes, and these accelerations will be those of a point having simple harmonic motions parallel to the axes. Since the constants which enter into the measure of the components of the force, and therefore into the measure of the accelerations produced by them, are the same for each component, the periods of these component simple harmonic motions will be the same. The motion of the particle will therefore be the resultant of two simple harmonic motions of equal periods at right angles to each other, and its path is therefore (§ 21) an ellipse, with its centre at the centre of force.

50. Central Force Proportional to the Inverse Square of the Radius Vector.—If the central force vary inversely with the square of the distance of the particle from the centre, the path described by the particle is in general a conic section, with the centre of force at one of its foci. To prove this we will use a theorem that will be demonstrated in § 55. It will there be shown that if a particle of mass μ be moved from an infinite distance under the action of a central force equal to $\frac{m}{r^2}$, where m is a constant and r the distance between the particle and the centre, the potential energy which it will lose by moving to a point distant r from the centre is given by $\frac{\mu m}{r}$. Thus, if μP represent the potential energy of the particle at an infinite distance, $\mu P - \frac{\mu m}{r}$ will represent its potential energy at the distance r. The sum of its potential and kinetic energies is constant; and hence $\mu P - \frac{\mu m}{r} + \frac{\mu v^2}{2} = \mu A$, a constant, or

$$\frac{v^2}{2} - \frac{m}{r} = A - P = C, \tag{34}$$

a constant. C may be greater or less than zero, or equal to zero.

If p_0 be the perpendicular let fall upon the line of direction of the moving particle at a chosen time when its velocity is v_0, and if p and v be the perpendicular and velocity at any other time, we know from § 48 that (35) $vp = v_0 p_0$. If we substitute $v = \frac{v_0 p_0}{p}$ in the above equation, we have $\frac{v_0^2 p_0^2}{2p^2} - \frac{m}{r} = C$, or

$$p^2 = \frac{v_0^2 p_0^2}{2} \frac{r}{m + Cr}. \tag{36}$$

This equation takes different forms depending upon the value of C. It becomes for

$$\left.\begin{aligned} C > 0,\ p^2 &= \frac{v_0^2 p_0^2}{2C} \cdot \frac{r}{\frac{m}{C} + r}; \\ C < 0,\ p^2 &= \frac{v_0^2 p_0^2}{2C} \cdot \frac{r}{\frac{m}{C} - r}; \\ C = 0,\ p^2 &= \frac{v_0^2 p_0^2}{2m} r. \end{aligned}\right\} \tag{37}$$

In these equations C has now a positive value. The first equation represents an hyperbola, the second an ellipse, and the third a parabola (Puckle's Conic Sections, §§ 204, 271). The focus of each of these conic sections is the point from which p and r are measured, or the centre of force.

The criteria which determine the nature of the curve may be otherwise given by $\frac{v^2}{2} > \frac{m}{r}$ for the hyperbola, $\frac{v^2}{2} < \frac{m}{r}$ for the ellipse, and $\frac{v^2}{2} = \frac{m}{r}$ for the parabola. That is, for the three curves respectively, the velocity of the particle at a point in its path is greater than, less than, or equal to the velocity which it would acquire by falling to that point from an infinite distance under the action of the central force.

The elements of the path may be obtained from these equations. The latus rectum of the parabola is $\frac{2v_0^2 p_0^2}{m}$. The semi-

major axis a and the semi-minor axis b of the hyperbola and ellipse are given by $2a = \frac{m}{C}$, $b^2 = \frac{v_0^2 p_0^2}{2C} = \frac{a v_0^2 p_0^2}{m}$. Hence $C = \frac{m}{2a}$, and using this value of C in the original equation, we get $\frac{v^2}{2} = \frac{m}{r} \pm \frac{m}{2a}$, the upper and lower signs holding for the hyperbola and ellipse respectively.

In case the particle is moving in an ellipse, its periodic time T, or the time in which it traverses the ellipse, may be found in terms of the elements of the ellipse and the constant m. The area of the ellipse is πab, and since the areas swept out in equal times by the radius vector drawn to the particle are equal, the rate at which the area is swept out is given by $\frac{\pi ab}{T}$. But $\frac{v_0 p_0}{2}$ also represents this rate, so that $\frac{\pi ab}{T} = \frac{v_0 p_0}{2}$. Substituting in this equation the value of $b = v_0 p_0 \left(\frac{a}{m}\right)^{\frac{1}{2}}$, we get (38) $T = \frac{2\pi a^{\frac{3}{2}}}{m^{\frac{1}{2}}}$, or, $m = \frac{4\pi^2 a^3}{T^2}$. If, therefore, different particles revolve in ellipses about a common centre of force in such a way that the squares of their periodic times are in the same ratio to the cubes of their semi-major axes, the constant m is the same for all of them.

51. The Problem of Two Bodies.—The problem of two bodies may be reduced to the problem of the action of a central force. For, suppose two particles to attract each other with a force given by $\frac{\mu m}{r^2}$, where μ and m are their masses and r the distance between them. The acceleration of the particle m, relative to the centre of mass, which will remain fixed in position, is given by $ma = \frac{\mu m}{r^2}$, or by $a = \frac{\mu}{r^2}$. The acceleration of the mass μ relative to the centre of mass is similarly $\frac{m}{r^2}$. If now an acceleration equal to $\frac{\mu}{r^2}$ and opposite to it in direction be impressed on both particles, the particle m will remain fixed, and the particle μ will move relatively to it

with the acceleration $\frac{\mu + m}{r^2}$. The path of the particle μ relative to the particle m will be therefore that due to a central force proceeding from m and equal to $\frac{\mu(\mu + m)}{r^2}$. The radius vector drawn to μ from m will still sweep out equal areas in equal times, and the path of μ will still be a conic section. If its path be an ellipse, the periodic time will be given by $T = \frac{2\pi a^{\frac{3}{2}}}{(\mu + m)^{\frac{1}{2}}}$; so that $\mu + m = \frac{4\pi^2 a^3}{T^2}$.

52. Motion of Projectiles.—In the special case of central motion in which the distance of the centre from the moving particle is very great and the velocity of the particle small, the particle describes a portion of an ellipse which differs very little from a parabola. This may be seen at once from the equation $\frac{v^2}{2} = \frac{m}{r} - \frac{m}{2a}$, for if r be very great and v small, the semi-major axis a must also be very great, and the path approaches the curve for which $2a$ is infinite, or a parabola.

This is the path described by a particle moving near the surface of the earth under the earth's attraction. The force which acts on the particle is really variable and directed toward the earth's centre, but within the limits of the path it may be considered constant and directed vertically downward.

This motion was first discussed by Galileo in connection with his study of falling bodies. His method was as follows: Let us assume the rectangular coordinates x and y, of which x is horizontal and y vertical, drawn upward in the direction opposite to the acting force. Let a particle be projected from the origin in the plane of the axes with the velocity v in a direction which makes with the x-axis the angle α.

The component velocities along the two axes are then $v \cos \alpha$ and $v \sin \alpha$. At the end of any time t reckoned from the instant at which the particle leaves the origin, the displacement of the

particle along the x-axis is $vt \cos \alpha$. If the force of gravity did not act on the particle its displacement along the y-axis in the same time would be $vt \sin \alpha$; but, since gravity acts, its real displacement along that axis is less than this by $s = \frac{1}{2}gt^2$, where g is the measure of the force or the acceleration of gravity, so that its displacement along the y-axis is $vt \sin \alpha - \frac{1}{2}gt^2$. The path of the particle, or the series of points which it occupies at successive instants, is found by eliminating t between the two equations for the two rectangular displacements. The equation of the path thus obtained is

$$y = \frac{\sin \alpha}{\cos \alpha} x - \frac{g}{2v^2 \cos^2 \alpha} x^2. \tag{39}$$

This represents a parabola passing through the origin. The axis is vertical, and the latus rectum is $\frac{2v^2 \cos^2 \alpha}{g}$. If $\alpha = 0$, or if the projection is horizontal, the equation becomes $x^2 = -\frac{2v^2}{g} y$, representing a parabola with its vertex at the origin.

When the body is projected above the horizontal plane, so that α lies between zero and $\frac{\pi}{2}$, it will attain its greatest height at the instant when its velocity along the y-axis becomes zero, or when $v \sin \alpha = gt$. The time required for it to describe its whole path and return to the x-axis is double this time or $\frac{2v \sin \alpha}{g}$. Its range, or the distance between its starting-point and the point at which it again meets the x-axis, is given by the product of this time and its horizontal velocity $v \cos \alpha$, or is $\frac{v^2}{g}\, 2 \sin \alpha \cos \alpha = \frac{v^2}{g} \sin 2\alpha$. The range is therefore a maximum when $\alpha = 45°$. Since $\sin(\pi - 2\alpha) = \sin 2\alpha$, the range is the same for projections at the angles α and $90° - \alpha$, or for projections equally inclined to the line bisecting the angle between the axes and on opposite sides of it.

53. Difference of Potential. The Potential.—Forces may arise from various causes. In any case they are only exhibited when

they affect the motion of bodies and they may be considered, for purposes of mathematical representation, as acting between the particles of the bodies. A *field of force* is a region in which a particle, constituting a part of a mutually interacting system, will be acted on by a force, and will move, if free to do so, in the direction of the force. The *strength of field* or the *intensity* in the field at a point is measured by the force which acts upon a unit quantity or *test unit* of that agent to which the force is due when placed at that point. The test unit is supposed not to affect the forces of the field.

When the force acting on a particle depends only on the position of the particle the study of its effects is often very much facilitated by the use of a concept called the *potential*. To explain this concept and its relation to force, we begin with a definition of the *difference of potential*. The difference of potential between two points in a field of force is equal to the work done by the forces of the field in moving a test unit from the one point to the other.

If $V_P - V_Q$ represent the difference of potential between the points P and Q, and if F represent the average force between those points and s the distance between them, then the amount of work done by that force in moving a test unit from P to Q, and hence the difference of potential between P and Q, is represented by

$$V_P - V_Q = Fs.$$

From this relation we have

$$F = \frac{V_P - V_Q}{s} = -\frac{V_Q - V_P}{s}. \tag{40}$$

If s become indefinitely small, in the limit F represents the force at the point P, and $-\dfrac{V_Q - V_P}{s} = -\dfrac{dV}{ds}$ becomes the rate of change of potential at that point with respect to space, taken with the opposite sign. Hence we obtain a definition of *potential*. It is a function, the rate of change of which at any point, with respect to space, taken with the opposite sign, measures the force at that point.

Let the test unit be situated at the point A and be moved over

any path to the point B. It is clear that if it be moved back over the same path from B to A the amount of work required to effect this motion will be equal and opposite to that done during the motion from A to B. This equality will also hold if the test unit be moved from B to A by any other path, provided the field of force is one which is nowhere interrupted by a region in which the force is not a function only of the position of the test unit, or is, as it is called, a *singly connected* region. The fields of force due to all forces known in Nature, except those caused by electrical currents, are singly connected regions. When the forces which act on the test unit at different points in its path are parallel, as in the case of gravity, this equality of the work done in carrying the test unit from one place to another over any path is obvious. If we assume the principle of the conservation of energy as a general principle, this equality may also be shown for fields in which the forces are not parallel; for, if the work done in moving the test unit over one path between A and B were not equal to that done in moving it over any other path between the same points, an endless supply of work could be obtained by repeatedly moving the unit over a path in which the work done by the forces of the field is greater, and returning it to its starting-point by motion over a path in which the work done is less. As this result is inconsistent with the principle of the conservation of energy, we conclude that the hypothesis from which it is deduced is untrue, and that the same amount of work will be done in moving the unit from the one point to the other, by whatever path the motion is effected. The difference of potential between the two points is therefore a function of their positions only.

54. Equipotential Surfaces and Lines of Force.—Let the test unit be moved from O along the different paths OA, OB, etc. (Fig. 22), so that the same amount of work is done upon it in each of these paths. The surface drawn through the end points of these paths is called an *equipotential surface ;* as may be seen from the proposition just proved, it is a surface in which the test unit may be moved without doing any work upon it. Since the forces

in the field will only do no work on the test unit when it is moved at right angles to their directions, it follows that the forces at the different points in an equipotential surface are normal to that surface. Draw the normals AP, BQ, etc., of such lengths that the work done in moving the test unit over them is the same. The surface drawn through their end points is again an equipotential surface. By repeating this process, the whole field of force may be mapped out by equipotential surfaces. In the limit, as the lengths of the normals thus drawn become infinitesimal, the successive normals will form continuous curves, everywhere normal to the equipotential surfaces which they cut. These curves, which represent the direction of the force at the points through which they pass, are called *lines of force.* If a small area be described on an equipotential surface, the lines of force which pass through its boundary will form a tubular surface, which will cut out corresponding areas on the other equipotential surfaces. This tubular surface, with the region enclosed by it, is called a *tube of force.*

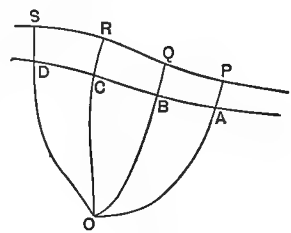

FIG. 22.

55. An Expression for Difference of Potential.—In what follows we will for convenience assume that the test unit is a unit mass, and that the field of force is due to the presence of particles which attract the unit mass with forces that are proportional to their masses and vary inversely with the square of the distance. By the proper choice of units the force due to any one particle may be set equal to $\frac{m}{r^2}$, where m is a constant proportional to the mass of the particle and r the distance between it and the test unit.

Let the point O (Fig. 23) be the point at which a particle m is placed, and let a unit mass traverse the path PRX under the action of the force $\frac{m}{r^2}$ directed toward O. When the particle is at

P, the force acting on it is $\frac{m}{OP^2}$; after it has moved through the infinitesimal distance PR, the force acting on it at R is $\frac{m}{OR^2}$. The work done upon it during this motion, equal to the product of the force and the line PQ, the projection of its path upon the direction of the force, is greater than $\frac{m \cdot PQ}{OP^2}$ and less than $\frac{m \cdot PQ}{OR^2}$; it may be shown that the work done during this displacement may be represented by $\frac{m \cdot PQ}{OP \cdot OR}$. Now $PQ = OP - OQ = OP - OR$ in the limit, so that the work done in this displacement equals $\frac{m(OP - OR)}{OP \cdot OR} = m\left(\frac{1}{OR} - \frac{1}{OP}\right)$. The work done in traversing the following elements of the path, RS, ST, etc., is expressed by $m\left(\frac{1}{OS} - \frac{1}{OR}\right)$, $m\left(\frac{1}{OT} - \frac{1}{OS}\right)$, etc. The work done in traversing the whole path from P to X is the sum of these expressions, or $m\left(\frac{1}{OX} - \frac{1}{OP}\right)$. By the definition of difference of potential, this expression is equal to the difference of potential between the points P and X, due to the force of which the centre is O. If the point X lie at an infinite distance from O, the work done by the force in moving the unit mass to that point equals $-\frac{m}{OP}$. This expression is called the *potential* at the point P. It has been obtained on the supposition that the force at P is directed toward O, or is an attractive force. In this case the test unit at an infinite distance possesses the potential energy E. In moving to P the forces of the field do upon it the work $\frac{m}{OP}$, so that its potential energy at P is $E - \frac{m}{OP}$.

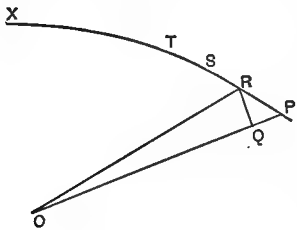

FIG. 23.

If the force at P and at the other points on the path be directed from O, the work done in the successive elements of the path is numerically equal to the expressions already obtained, but is opposite in sign; so that the work done by such forces, as the test unit moves from P to X, is equal to $m\left(\frac{1}{OP} - \frac{1}{OX}\right)$. When the point X is at an infinite distance from O, the work done in moving the test unit from P to X equals $\frac{m}{OP}$. This is the potential at the point P, due to a repulsive force with its centre at O. In this case the test unit at an infinite distance has no potential energy, so that $\frac{m}{OP}$ expresses its potential energy at P.

56. Flux of Force. Tubes of Force.—Still retaining the convention that the forces of the field are due to mass attraction and follow the law of inverse squares, we will now prove certain propositions which are of great importance in the theories of gravitation, electricity, and magnetism.

If in a field an area s be described so small that the force is the same for all points of it, the product of the area and the normal component of the force is called the elementary *flux of force* over or through that area. We will show that the total flux of force, that is, the sum of all the elementary fluxes, taken over a closed surface in the field which does not contain any masses is equal to zero.

We consider first the flux of force arising from a mass m situated at the point O. Let ABC (Fig. 24) represent a closed surface not containing the mass m; draw a tube of force cutting this surface in the elements s and s' The forces due to the mass m at points in these areas will be $\frac{m}{r^2}$ and $\frac{m}{r'^2}$ respectively.

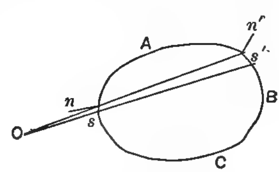

Fig. 24.

We represent the angles between the common direction of these forces and the normals to the elements s and s'

drawn outward from the surface by α and α' respectively. The components of the forces $\frac{m}{r^2}$ and $\frac{m}{r'^2}$ drawn outward normal to the surfaces s and s' are $\frac{m}{r^2}\cos\alpha$ and $\frac{m}{r'^2}\cos\alpha'$ respectively. Hence the flux of force through these elements is $\frac{m}{r^2}s\cos\alpha + \frac{m}{r'^2}s'\cos\alpha'$. But $s\cos\alpha$ and $-s'\cos\alpha'$ are equal to the normal cross-sections of the tube of force at the distances r and r' from O, the minus sign being inserted because one of the two cosines is negative; and since the tube of force is a cone,

$$\frac{s\cos\alpha}{r^2} = -\frac{s'\cos\alpha'}{r'^2}.$$

Hence the flux of force through these two elements, due to the mass at the point O, is equal to zero. Since similar tubes of force may be drawn from the point O so as to include all the elements of the surface ABC, and since to each pair of elements thus determined the same proposition applies, it follows that the total flux of force due to the mass m through the surface is equal to zero. The same proposition will hold for the flux of force due to any other particle situated outside the surface, and therefore holds true for any mass whatever situated outside the surface.

The flux of force through a closed surface containing any number of particles is equal to $4\pi M$, where M is the mass of all the particles. To prove this, let us consider a single particle m situated at the point O. About this point describe a sphere of radius r. The force at each point of the sphere is $\frac{m}{r^2}$, and the total flux of force through the sphere is equal to this force multiplied by the area of the sphere, or to $\frac{m}{r^2}.4\pi r^2 = 4\pi m$. Now to prove a similar proposition for any closed surface enclosing the mass m, we describe about the point O a sphere which is entirely enclosed by the surface. Since the region enclosed between this sphere and

the surface contains no mass, the total flux of force through it equals zero. But the flux of force through the sphere equals $-4\pi m$, the minus sign being used because the normals to the sphere, when considered as bounding the region enclosing the mass and as bounding the region between it and the closed surface, have opposite directions. Therefore the flux of force through the closed surface must equal $4\pi m$. This proposition holds for each of the masses contained within the closed surface, so that, if the sum of these masses be M, the total flux of force through the closed surface is $4\pi M$.

Let us apply the theorem just proved to the region bounded by a tube of force of very small cross-section and by two rectangular cross-sections of the tube. Since the tube of force is everywhere bounded by lines of force, and since, therefore, the force at a point on the tube has no component normal to the surface of the tube, the only parts of the closed surface under consideration which contribute to the flux of force through it are the two end cross-sections. Represent the areas of the two cross-sections by s and s', and the forces acting at them by F and F' respectively. Then, since the total flux of force equals zero, we have $Fs + F's' = 0$ or $Fs = -F's'$. The minus sign appears because the force and the normal to the cross-section are in the same direction at one end of the tube and in opposite directions at the other. If we confine our attention to the numerical value of the product Fs, we may say that the flux of force is the same for all cross-sections of the tube of force. This proposition, though here proved only for a tube of force of very small cross-section, manifestly may be generalized for any tube of force whatever.

57. Special Cases.—It is sometimes important, especially in the study of electricity, to know the force which is exerted by a plane sheet of matter at a point near it. We call the quantity of matter which is enclosed by a unit area drawn on such a sheet the *surface density* of the sheet at the point where the area is taken; more strictly, the surface density is the ratio of the quantity of matter enclosed by the area to the magnitude of the area, as the area di-

minishes indefinitely. If we consider *an infinitely extended plane sheet*, it is evident that the lines of force in the region near it are perpendicular to its surface. Take any small area on the surface of the sheet, and consider the closed surface bounded by the lines of force which pass through the boundary of that area and by two cross-sections taken parallel with the sheet on the opposite sides of it. The flux of force through the sides of the surface thus formed is zero, because the lines of force lie in that surface. The only portions of the surface, therefore, which contribute to the flux of force, are the end cross-sections. Let s represent the area of each of these cross-sections, which are equal, F the force at one of them, and F' that at the other. If σ represent the surface density of the sheet, σs is the mass enclosed within the closed surface. Applying the theorem of the flux of force, we have $(F + F')s = 4\pi\sigma s$ or $F + F' = 4\pi\sigma$. Remembering that the directions of these forces are outward from the closed surface, and that therefore $+ F$ and $- F'$ are forces drawn in the same direction along the lines of force, this equation shows that in passing through a sheet of surface density σ the force changes by $4\pi\sigma$. If the forces in the field be due only to the sheet, it is manifest, from symmetry, that the force F and the force F' are equal, and that their directions are opposite. We thus have $F + F' = 2F = 4\pi\sigma$, or $F = 2\pi\sigma$. That is, the force at a point infinitely near a plane sheet of surface density σ is equal to $2\pi\sigma$. This proposition holds, even if the sheet be not plane, for any points so near it that the neighboring lines of force are parallel.

The force *within a closed spherical shell* of uniform surface density vanishes at every point. For, let us construct a sphere in the region contained by the shell and concentric with it. Since no matter is contained by this sphere, the total flux of force through its surface is zero, and since, by symmetry, the force at any point on the inner sphere must have the same value and the same direction to or from the centre, it follows that $\Sigma Fs = F4\pi r^2 = 0$, and hence that $F = 0$. The force, therefore, vanishes for all points in the interior of the shell. It manifestly vanishes also within a

closed surface formed of concentric spherical shells, in each one of which the surface density is uniform.

The force at a point *outside a spherical shell* of uniform surface density varies inversely with the square of the distance between that point and the centre of the spherical shell. For, describe a sphere concentric with the shell and of radius r, greater than the radius of the shell. Applying to this sphere the theorem of the flux of force, we have $\Sigma Fs = 4\pi M$, where M is the mass of the spherical shell. It is evident, by symmetry, that the force at every point on the sphere to which this theorem is applied must be the same in magnitude and similarly directed along the radius of the sphere. The flux of force ΣFs therefore equals $F \cdot 4\pi r^2 = 4\pi M$, or $F = \frac{M}{r^2}$. This theorem manifestly holds also for the force at a point outside any mass bounded by a spherical surface, provided that the matter in the sphere is distributed uniformly or in concentric shells, in each one of which the surface density is uniform.

CHAPTER II.

MASS ATTRACTION.

58. Mass Attraction.—The law of *mass attraction* was the first generalization of modern science. It may be stated as follows:—

Between every two material particles in the universe there is a stress, tending to move them toward each other, which varies directly as the product of the masses of the particles, and inversely as the square of the distance between them. This law is sometimes called the law of *universal attraction* and sometimes the law of *gravitation.*

Some of the ancient philosophers had a vague belief in the existence of an attraction between the particles of matter. This hypothesis, however, with the knowledge which they possessed, could not be proved. The geocentric theory of the planetary system, which obtained almost universal acceptance, offered none of those simple relations of the planetary motions upon which the law was finally established. It was not until the heliocentric theory of Copernicus had been established by the discoveries of Galileo, and the labors of Kepler, that the discovery of the law became possible.

In particular, the three laws of planetary motion published by Kepler in 1609 and 1619 laid the foundation for Newton's demonstrations. The laws are as follows:—

I. The planets move in ellipses of which one focus is situated as the sun.

II. The radius vector drawn from the sun to the planet sweeps out equal areas in equal times.

III. The squares of the periodic times of the planets are proportional to the cubes of the semi-major axes of their orbits.

Kepler could give no physical reason for the existence of such laws. Later in the century, after Huygens had discovered certain theorems relating to motion in a circle, it was seen that the third law would hold true for bodies moving in concentric circles, and attracted to the common centre by forces varying inversely as the squares of the radii of the circles. Several English philosophers, among them Hooke, Wren, and Halley, based a belief in the existence of an attraction between the sun and the planets upon this theorem.

The demonstration was by no means a rigorous one, and was not generally accepted. It was left for Newton to show that not only the third, but all, of Kepler's laws were completely satisfied by the assumption of the existence of an attraction acting between the sun and the planets, varying inversely as the square of the distance. The demonstrations which show that the law of universal attraction is consistent with Kepler's laws are given in §§ 48, 50.

Newton also showed that the attraction holding the moon in its orbit, which is presumably of the same nature as that existing between the sun and the planets, is of the same nature as that which causes heavy bodies to fall to the earth. This he accomplished by showing that the deviation of the moon from a rectilinear path is such as should occur if the force which at the earth's surface is the *force of gravity* were to extend outwards to the moon, and vary in intensity inversely as the square of the distance.

Two further steps were necessary before the final generalization could be reached. One was, to show the relation of the attraction to the masses of the attracting bodies; the other, to show that this attraction exists between all particles of matter, and not merely, as Huygens believed, between those particles and the centres of the sun and planets.

The first step was taken by Newton. By means of pendulums having the same length, but with bobs of different materials, he showed that the force acting on a body at the earth's surface is

proportional to the mass of the body, since all bodies have the same acceleration. He further brought forward, as the most satisfactory theory which he could form, the general statement that every particle of matter attracts and is attracted by every other particle.

The experiments necessary for a complete verification of this last statement were not carried out by Newton. They were performed in 1798 by Cavendish. His apparatus consisted essentially of a bar furnished at both ends with small leaden balls, suspended horizontally by a long fine wire, so that it turned freely in the horizontal plane. Two large leaden balls were mounted on a bar of the same length, which turned about a vertical axis coincident with the axis of rotation of the suspended bar. The large balls, therefore, could be set and clamped at any angular distance desired from the small balls. The whole arrangement was enclosed in a room, to prevent all disturbance. The motion of the suspended system was observed from without by means of a telescope. Neglecting as unessential the special methods of observation employed, it is sufficient to state that an attraction was observed between the large and small balls, and was found to be in accordance with the law as above stated.

59. Centre of Gravity.—The forces with which the earth attracts the particles of an ordinary body are parallel and proportional to the masses of the particles, so that the sum of their moments about any axis passing through the centre of mass will vanish, because the corresponding sum of the products of the masses and their respective distances from any plane containing that axis vanishes by the definition of the centre of mass. Gravity will, therefore, have no tendency to produce rotation in a free body or system of particles. It will cause a translation of the body, if it be rigid, such as would be produced if a force equal to the sum of all the forces acting on the particles were applied at the centre of mass. This point of application of the force is called the *centre of gravity* of the body. If the forces acting on the particles be not parallel, the body will, in general, have no centre of gravity. Cer-

tain bodies, which have a centre of gravity even when the forces are not parallel, are called *centrobaric* bodies.

60. Measurement of the Force of Gravity.—When two bodies attract each other, their accelerations relative to their fixed centre of mass are inversely as their masses. In the case of the attraction between the earth and a body near its surface, the mass of the earth is so great that its acceleration may be neglected and the acceleration of the body alone need be considered. Since the force acting upon it varies with its mass, and since its gain in momentum also varies with its mass, it follows that its acceleration will be the same, whatever its mass may be. We may, therefore, obtain a direct measure of the earth's attraction, or of the *force of gravity*, by allowing a body to fall freely, and determining its acceleration. It is found that a body so falling at latitude 40° will describe in one second about 16.08 feet, or 490 centimetres. Its acceleration is therefore 32.16 in feet and seconds or 980 in centimetres and seconds. We denote this acceleration by the symbol g.

The force acting on the body, or the *weight* of the body, is seen at once to be mg, where m is the mass of the body.

On account of the difficulties in the employment of this method, various others are used to obtain the value of g indirectly. For example, we may allow bodies to slide or roll down a smooth inclined plane, and observe their motion. The force effective in producing motion on the plane is $g \sin \phi$, where ϕ is the angle of the plane with the horizontal; the space traversed in the time t is $s = \frac{1}{2}gt^2 \sin \phi$. By observing s and t, the value of g may be obtained. The motion is so much less rapid than that of a freely falling body that tolerably accurate observations can be made. Irregularities due to friction upon the plane and the resistance of the air, however, greatly vitiate any calculations based upon these observations. This method was used by Galileo in his investigation of the laws of falling bodies.

The most exact method for determining the value of g is based upon observations of the oscillation of a pendulum.

A *pendulum* may be defined as a heavy mass, or bob, suspended

from a rigid support, so that it can oscillate about its position of equilibrium.

FIG. 25.

In the *simple*, or mathematical, pendulum the bob is assumed to be a material particle, and to be suspended by a thread without weight. If the bob be stationary and acted on by gravity alone, the line of the thread will be the direction of the force. If the bob be withdrawn from the position of equilibrium (Fig. 25), it will be acted on by a force at right angles to the thread, in a direction opposite that of the displacement, expressed by $g \sin \phi$, where ϕ is the angle between the perpendicular and the new position of the thread.

The force acting upon the bob at any point in the circle of which the thread is radius, if it be released and allowed to swing in that circle, varies as the sine of the angle between the perpendicular and the radius drawn to that point. If we make the oscillation so small that the arc may be substituted for its sine without sensible error, the force acting on the bob varies as the displacement of the bob from the point of equilibrium.

A body acted on by a force varying as the displacement of the body from a fixed point will have a simple harmonic motion about its position of equilibrium (§ 21).

Hence it follows that the oscillations of the pendulum are symmetrical about the position of equilibrium. The bob will have an amplitude on the one side of the vertical equal to that which it has on the other, and the oscillation, once set up, will continue forever unless modified by outside forces.

The importance of the pendulum as a means of determining the value of g consists in this: that, instead of observing the space traversed by the bob in one second, we may observe the number of oscillations made in any period of time, and determine the time of one oscillation; from this, and the length of the pendulum, we can calculate the value of g. The errors in the necessary observations and measurements are very slight in comparison with those of any other method.

61. Formula for Simple Pendulum.—The formula connecting the time of oscillation with the value of g is obtained as follows: The acceleration of the bob at any point in the arc is, as we have seen, $g \sin \phi$, or $g\phi$ if the arc be very small. The acceleration in a simple harmonic motion is $-\omega^2 s = -\frac{4\pi^2}{T^2}s$, where s is the displacement.

Since the bob has a simple harmonic motion, we may set these two expressions for the acceleration equal, neglecting the minus sign, which merely expresses the fact that the acceleration is toward the centre of the path; hence $g\phi = \frac{4\pi^2}{T^2}s$.

The displacement s is equal to $l\phi$, if l represent the length of the thread; hence $g = \frac{4\pi^2 l}{T^2}$, from which $T = 2\pi\sqrt{\frac{l}{g}}$.

In this formula T represents the time of a double oscillation. It is customary to observe the time of a single oscillation, when the formula becomes

$$t = \pi\sqrt{\frac{l}{g}}. \qquad (41)$$

62. Physical Pendulum.—Any pendulum fulfilling the requirements of the foregoing theory is, of course, unattainable in practice. We may, however, calculate from the known dimensions and mass of the portions of matter making up the *physical pendulum*, what would be the length of a simple pendulum which would oscillate in the same time. It is clear that there must be some point in every physical pendulum the distance of which from the point of suspension is equal to the length of the corresponding simple pendulum; for the particles near the point of suspension tend to oscillate more rapidly than those more remote, and the time of oscillation of the system, if it be rigid, will be intermediate between the times of oscillation which the particles nearest to, and most remote from the point of suspension would have if they were oscillating freely. There will, therefore, be some one particle of which the proper rate

of oscillation is the same as that of the whole pendulum. Its distance from the point of suspension is the length sought.

In determinations of the value of g by observations upon the time of oscillation of a pendulum, the length of the equivalent simple pendulum may be found in either of two ways:

(1) The pendulum may be constructed in such a manner that its moment of inertia and the position of its centre of gravity may be calculated. From these data the required length is readily obtained.

To show this, we consider any mass swinging as a pendulum about a horizontal axis. The force which sets it in oscillation is its weight Mg. The effect of this force in producing rotation about the axis is given by $MgR \sin \phi = I\alpha$ (§ 39), where I is its moment of inertia about that axis and R is the distance from the axis to its centre of gravity. As in the case of the simple pendulum, when the oscillations are infinitesimal, $\sin \phi$ may be replaced by ϕ. Now ϕ and α represent the angular displacement and the angular acceleration of any point of the pendulum, and the actual displacement and acceleration are proportional to them; and since the displacement and acceleration are proportional to each other, every point in the pendulum has a simple harmonic motion of the same period. The actual acceleration of the centre of mass equals $R\alpha = \frac{MgR}{I} \cdot R\phi$. Now $R\phi$ is the displacement of the centre of mass, and therefore from the formula connecting acceleration and displacement in simple harmonic motion, used in § 61, we obtain $\frac{4\pi^2}{T^2} = \frac{MgR}{I}$. Hence $T = 2\pi\sqrt{\frac{I}{MRg}}$. Or, if we designate by t the time of oscillation from one extremity of the arc to the other, we have

$$t = \pi\sqrt{\frac{I}{MRg}}. \tag{42}$$

We may replace I by its equivalent $I' + MR^2$, where I' is the moment of inertia about an axis parallel to the axis of suspension and passing through the centre of gravity. By comparison of this

equation with the one giving the time of oscillation of a simple pendulum, it appears that the length l of the simple pendulum which will oscillate in the same time as the physical pendulum, or, as it is called, the length of the *equivalent simple pendulum*, is given by

$$l = \frac{I}{MR} = \frac{I' + MR^2}{MR}. \tag{43}$$

A line drawn parallel with the axis of suspension, through a point at the distance l from that axis and on the line drawn through the centre of gravity perpendicular to that axis, is called the *axis of oscillation*. It evidently contains the centre of percussion (§ 44).

A pendulum consisting of a heavy spherical bob suspended by a cylindrical wire was used by Borda in his determinations of the value of g. The moment of inertia and the centre of gravity of the system were easily calculated, and the length of the simple pendulum to which the system was equivalent was thus obtained.

(2) We may determine the length of the equivalent simple pendulum directly by observation. The method depends upon the principle that, if the axis of oscillation be taken as the axis of suspension, the time of oscillation will not vary. The proof of this principle is as follows:

Suppose the pendulum suspended so as to swing about the axis of oscillation as a new axis of suspension. The distance of the axis of oscillation from the centre of gravity is $l - R$, and the length l' of the equivalent simple pendulum, in this case, is $l' = \dfrac{I' + M(l-R)^2}{M(l-R)}$. Now $l = \dfrac{I' + MR^2}{MR}$ or $I' = MR\,(l - R)$, and substituting this value in the equation for l' and reducing, we obtain $l' = l$. That is, the length of the equivalent simple pendulum, and consequently the time of oscillation when the pendulum swings about its axis of suspension, is the same as that when it is reversed and swings about its former axis of oscillation.

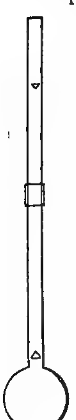

Fig. 26.

A pendulum (Fig. 26) so constructed as to take advantage of this principle was used by Kater in his determination of the value of g; and this form is known, in consequence, as Kater's pendulum.

63. The Balance.—The weights of bodies, and hence also their masses, are compared by means of the *balance.*

To be of value, the balance must be accurate and sensitive; that is, it must be in the position of equilibrium when the scale-pans contain equal masses, and it must move out of that position on the addition to the mass in one pan of a very small fraction of the original load.

The balance consists essentially of a regularly formed beam, poised at the middle point of its length upon knife-edges which rest on agate planes. From each end of the beam is hung a scale-

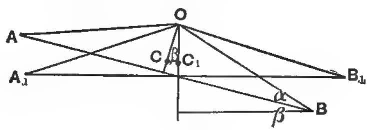

FIG. 27.

pan, in which the masses to be compared are placed. Let O (Fig. 27) be the point of suspension of the beam; A, B, the points of suspension of the scale-pans; C, the centre of gravity of the beam, the weight of which is W. Represent $OA = OB$ by l, OC by d, and the angle OAB by α.

If the weight in the scale-pan at A be P, and that in the one at B be $P + p$, where p is a small additional weight, the beam will turn out of its original horizontal position, and assume a new one. Let the angle COC'', through which it turns, be designated by β. Then the moments of force about O are equal; that is,

$$(P + p)l \,.\, \cos(\alpha + \beta) = Pl \,.\, \cos(\alpha - \beta) + Wd \,.\, \sin\beta;$$

from which we obtain, by expanding and transposing,

$$\tan\beta = \frac{pl\cos\alpha}{(2P + p)l\sin\alpha + Wd}. \tag{44}$$

The conditions of greatest sensitiveness are readily deducible from this equation. So long as $\cos\alpha$ is less than unity, it is evident that $\tan\beta$, and therefore β, decreases as the weight $2P$ of

the load increases. As the angle α becomes less, the value of β increases for a given load, and is less affected by changes in the load, until, when A, O, and B are in the same straight line, it depends only on $\frac{pl}{Wd}$, and is independent of the load. In this case $\tan\beta$ increases as d, the distance from the point of suspension to the centre of gravity of the beam, diminishes, and as the weight of the beam W diminishes. To secure sensitiveness, therefore, the beam must be as long and as light as is consistent with stiffness, the points of suspension of the beam and of the scale-pans must be very nearly in the same line, and the distance of the centre of gravity from the point of suspension of the beam must be as small as possible. Great length of beam, and near coincidence of the centre of gravity with the axis, are, however, inconsistent with rapidity of action. The purpose for which the balance is to be used must determine the extent to which these conditions of sensitiveness shall be carried.

Accuracy is secured by making the arms of the beam of equal length, and so that they will perfectly balance, and by attaching scale-pans of equal weight at equal distances from the centre of the beam.

In the balances usually employed in physical and chemical investigations, various means of adjustment are provided, by means of which all the required conditions may be secured. The beam is poised on knife-edges; and the adjustment of its centre of gravity is made by changing the position of a nut which moves on a screw, placed vertically, directly above the point of suspension. Perfect equality in the moments of force due to the two arms of the beam is secured by a similar horizontal screw and nut placed at one end of the beam. The beam is a flat rhombus of brass, large portions of which are cut out so as to make it as light as possible. The knife-edge on which the beam rests, and those upon which the scale-pans hang, are arranged so that, with a medium load, they are all nearly in the same line. A long pointer attached to the beam moves before a scale, and serves to indicate the deviation of the

beam from the position of equilibrium. If the balance be accurately made and perfectly adjusted, and equal weights placed in the scale-pans, the pointer will remain at rest, or will oscillate through distances regularly diminishing on each side of the zero of the scale.

If the weight of a body is to be determined, it is placed in one scale-pan, and known weights are placed in the other until the balance is in equilibrium, or nearly so. The final determination of the exact weight of the body is then made by one of three methods: we may continue to add very small weights until equilibrium is established; or we may observe the deviation of the pointer from the zero of the scale, and, by a table prepared empirically, determine the excess of one weight over the other; or we may place a known weight at such a point on a graduated bar attached to the beam that equilibrium is established, and find what its value is, in terms of weight placed in the scale-pan, by the relation between the length of the arm of the beam and the distance of the weight from the middle point of the beam.

If the balance be not accurately constructed, we can, nevertheless, obtain an accurate value of the weight desired. The method employed is known as Borda's method of double weighing. The body to be weighed is placed in one scale-pan, and balanced with fine shot or sand placed in the other. It is then replaced by known weights till equilibrium is again established. It is manifest that the replacing weights represent the weight of the body.

If the error of the balance consist in the unequal length of the arms of the beam, the true weight of a body may be obtained by weighing it first in one scale-pan and then in the other. The geometrical mean of the two values is the true weight; for let l_1 and l_2 represent the lengths of the two arms of the balance, P the true weight, and P_1 and P_2 the values of the weights placed in the pans at the extremities of the arms of lengths l_1 and l_2, which balance it. Then $Pl_2 = P_1 l_1$ and $Pl_1 = P_2 l_2$; from which $P = \sqrt{P_1 P_2}$.

64. Density of the Earth. Constant of Mass Attraction.—One of

the most interesting problems connected with the physical aspect of gravitation is the determination of the constant of mass attraction. It has been attacked in several ways, each of which is worthy ot consideration. The methods employed usually depend upon a determination of the mean density of the earth.

The first successful determination of the earth's density was based upon experiments made in 1774 by Maskelyne. He observed the deflection from the vertical of a plumb-line suspended near the mountain Schehallion in Scotland. He then determined the density of the mountain by the specific gravity of specimens of earth and rock from various parts of it, and calculated the ratio of the volume of the mountain to that of the earth. From these data the mean density of the earth was determined to be about 4.7.

The next results were obtained from the experiments of Cavendish, in 1798, with the torsion balance already described. The density, volume, and attraction of the leaden balls being known, the constant of mass attraction could be calculated, and also the density of the earth obtained. The value obtained by Cavendish for the latter was about 5.5.

Another method, employed by Carlini in 1824, depends upon the use of the pendulum. The time of the oscillation of a pendulum at the sea-level being known, the pendulum is carried to the top of some high mountain, and its time of oscillation again observed. The value of g as deduced from this observation will, of course, be less than that obtained by the observation at the sea-level. It will not, however, be as much less as it would be if the change depended only on the increased distance from the centre of the earth. The discrepancy is due to the attraction of the mountain, which can, therefore, be calculated, and the calculations completed as in Maskelyne's experiment. The value obtained by Carlini by this method was about 4.8.

A fourth method, due to Airy, and employed by him in 1854, consists in observing the time of oscillation of a pendulum at the bottom of a deep mine. By § 57 it appears that the attraction of a spherical shell of earth the thickness of which is the depth of the

mine vanishes. The mean density of the earth may, therefore, be determined by the discrepancy between the values of g at the bottom of the mine and at the surface.

Still another method, used by Jolly, consists in determining by means of a delicate balance the increase in weight of a small mass of lead when a large leaden block is brought beneath it. Jolly's results were very consistent, and give as the earth's density the value 5.69.

These methods have yielded results varying from that obtained by Airy, who stated the mean specific gravity to be 6.623, to that of Maskelyne, who obtained 4.7. The most elaborate experiments, by Cornu and Baille, by the method of Cavendish, gave as the value 5.56. This is probably not far from the truth.

. When the density of the earth is known, we may calculate from it the value of the *constant of mass attraction*, that is, the attraction between two unit masses at unit distance apart. Represent by D the earth's mean density, by R the earth's mean radius, and by k the constant of attraction. The mass of the earth is expressed by $\frac{4}{3}\pi R^3 D$. Since by § 57 the attraction of a sphere is inversely as the square of the distance from its centre, the attraction of the earth on a gram at a point on its surface, or the weight of one gram, is expressed by $g = \frac{4}{3}\pi\frac{R^3 D}{R^2}k = \frac{4}{3}\pi RDk$. πR is twice the length of the earth's quadrant, or 2×10^9 centimetres. The value of g at latitude 40° is 980.11, and from the results of Cornu and Baille we may set D equal to 5.56. With these data we obtain k equal to 0.000000066 dynes.

CHAPTER III.

MOLECULAR MECHANICS.

CONSTITUTION OF MATTER.

65. General Properties of Bodies. — Besides the properties already defined in § 3 as characteristic and essential properties of matter, we find that all bodies possess the properties of compressibility and divisibility.

Compressibility.—All bodies change in volume by change of pressure and temperature. If a body of a given volume be subjected to pressure it will return to its original volume when the pressure is removed, provided the pressure has not been too great. This property of assuming its original volume is called *elasticity.* The property of changing volume by the application of heat is sometimes specially called *dilatability.*

Divisibility.—Any body of sensible magnitude may, by mechanical means, be divided, and each of its parts may again be subdivided; and the process may be continued till the resulting particles become so minute that we are no longer able to recognize them, even when assisted by the most perfect appliances of the microscope. If the body be one that can be dissolved, it may be put in solution, and this may be greatly diluted; and in some cases the body may be detected by the color which it imparts to the diluent, even when constituting so small a proportion as one one-hundred-millionth part of the solution.

66. Molecules.—We are not, however, at liberty to conclude that matter is infinitely divisible. The fact, established by observation, that bodies are impenetrable, and the one just noted, that they are also compressible, as well as other considerations, to be adduced later, lead to the opposite conclusion. To explain the

coexistence of these properties we are compelled to assume that bodies are composed of extremely small portions of matter, indivisible without destroying their identity, called *molecules*, and that these molecules are separated by interstitial spaces occupied by a medium called the *ether*.

These molecules can be divided only by chemical means. The resulting subdivisions are called *atoms*. The atom, however, cannot exist in a free state. The molecule is the physical unit of matter, while the atom is the chemical unit.

67. Composition of Bodies.—It has just been said that atoms cannot exist in a free state. They are always combined with others, either of the same kind, forming simple substances, or of dissimilar kinds, forming compound substances.

There are about seventy substances now known which cannot, in the present state of our knowledge, be decomposed, or made to yield anything simpler than themselves. We therefore call them *simple substances*, *elements*, or, if we desire to avoid expressing any theory concerning them, *radicals*. It is not improbable that some of these will yet be divided, perhaps all of them. We can call them elements, then, only provisionally.

68. States of Aggregation.—Bodies exist in three states—the solid, the liquid, and the gaseous. In the *solid* state the form and volume of the body are both definite. In the *liquid* state the volume only is definite. In the *gaseous* state neither form nor volume is definite.

Many substances may, under proper conditions, assume either of these three states of aggregation; and some substances, as, for example, water, may exist in the three states under the same general conditions.

It is proper to add, however, that there is no such sharp line of distinction between the three states of matter as our definitions imply. Bodies present all gradations of aggregation between the extreme conditions of solid and gas; and the same substance, in passing from one state to the other, often presents all these gradations.

69. Structure of Solids.—With the exception of organized bodies, all solids may be divided into two classes. The bodies of one class, which are characterized by more or less regularity of form, are called *crystalline;* those of the other class, exhibiting no such regularity, are called *amorphous.* For the formation of crystals a certain amount of freedom of motion of the molecules is necessary. Such freedom of motion is found in the gaseous and liquid states; and when crystallizable bodies pass slowly from these to the solid state, crystallization usually occurs. It may also occur in some solids spontaneously, or in consequence of agitation of the molecules by mechanical means, such as friction or percussion. Crystallizable bodies are called *crystalloids.*

Some amorphous bodies cannot, under any circumstances, assume the crystalline form. They are called *colloids.*

70. Crystal Systems.—Crystals are arranged by mineralogists in six systems.

In the first, or *Isometric,* system all the forms are referred to three equal axes at right angles. The system includes the cube, the regular octahedron, and the rhombic dodecahedron.

In the second, or *Dimetric,* system all the forms are referred to a system of three rectangular axes, of which only two are equal.

In the third, or *Hexagonal,* system the forms are referred to four axes, of which three are equal, lie in one plane, and cross each other at angles of 60°. The fourth axis is at right angles to the plane of the other three, and passes through their common intersection.

The fourth, or *Orthorhombic,* system is characterized by three rectangular axes of unequal length.

In the fifth, or *Monoclinic,* system the three axes are unequal. One of them is at right angles to the plane of the other two. The angles which these two make with each other, as well as the relative lengths of the axes, vary greatly for different substances.

In the sixth, or *Triclinic,* system the three axes are oblique to each other, and unequal in length.

71. Forces Determining the Structure of Bodies.—In view of

what precedes, it is necessary to assume the existence of certain forces other than the mass attraction considered in § 58, acting between the molecules of matter. These forces seem to act only within very small or insensible distances, and vary with the character of the molecule. They are hence called *molecular forces.* In liquids and solids there must be a force of the nature of attraction holding the molecules together, and a force equivalent to repulsion preventing actual contact. The attractive force is called *cohesion* when it unites molecules of the same kind, and *adhesion* when it unites molecules of different kinds. The repulsive force is probably a manifestation of that motion of the molecules which constitutes heat. In gases this motion is so great as to carry the molecules beyond the limit of their mutual molecular attractions: thus the apparent repulsion prevails, and the gas only ceases expanding when this repulsion is balanced by other forces.

72. Structure of the Molecule.—The facts brought to light in the study of crystals compel us to ascribe a structural form to the molecule, determining special points of application for the molecular forces. From this results the arrangement of molecules which have the requisite freedom of motion into regular crystalline forms.

73. Nature of the Atom.—The *atom,* or the least part into which matter can be divided by any means now known, must itself possess inertia and impenetrability. Our inability to divide the atom, and the demonstration by Lavoisier and others that none of the matter which takes part in a chemical change is destroyed by that change, lead us to assert that the atom is also indestructible. The kinetic theory of heat requires the additional assumption that the atom is generally in motion; and the existence of molecular forces and of chemical combination lead us to assert also that the atoms exert force on one another. These properties were summed up by Newton, who first gave a description of the atom, in a form suitable for use in physical science, in the following words: "It seems probable to me that God in the beginning formed matter in solid, massy, hard, impenetrable, movable

particles, of such sizes and figures and with such other properties and in such proportion to space as most conduced to the end for which He formed them; and that these primitive particles, being solids, are incomparably harder than any porous bodies compounded of them, even so very hard as never to wear or break in pieces; no ordinary power being able to divide what God Himself made one in the first creation. . . . It seems to me, farther, that these particles have not only a *vis inertiæ,* accompanied with such passive laws of motion as naturally result from that force, but also that they are moved by certain active principles."

The science of physics has been erected upon the conception of the atom embodied in this description. It is not, however, a theory of the nature of the atom in any proper sense: any theory must explain the various properties ascribed to the atom on mechanical principles, with as few assumptions as are necessary to apply those principles. Such a theory was proposed by Thomson: it is known as the *vortex atom theory.* This theory assumes that the space occupied by the universe is filled with a continuous, incompressible, perfect fluid, and that each atom is a small closed vortex in this fluid. A comparison of the properties of the atom with those of the closed vortex, described in § 122, shows that the two sets of properties are identical. Thus the vortex, like the atom, retains its identity throughout any changes it may undergo, that is, it is not composed merely of matter in similar states of motion, but of identically the same matter at all times. No two vortices can cut each other or can occupy the same space at the same time; they are therefore indestructible and impenetrable. Furthermore, a vortex must move as a whole, and any two vortices near each other change their directions of motion as if they exerted force on each other. A vortex in which the vortex line is implicated or knotted any number of times will always retain the same degree of implication; so that if a vortex of a special sort be once set up, it will always retain its essential characteristics, corresponding to the retention of special characteristics by the atoms of the different elements. This theory, taken in connection with Fitz-

gerald's theory of the vortex ether (§ 323), gives an almost complete model of the essential features of the physical universe; it does not, however, explain gravitation, nor does it without some addition explain the inertia of a body, and not until it is shown that these characteristic features of matter are explained by it can it be adopted as a final theory of matter.

FRICTION.

74. General Statements.—When the surface of one body is made to move over the surface of another, a resistance to the motion is set up. This resistance is said to be due to *friction* between the two bodies. It is most marked when the surfaces of two solids move over one another. It exists, however, also between the surfaces of a solid and of a liquid or a gas, and between the surfaces of contiguous liquids or gases. When the parts of a body move among themselves, there is a similar resistance to the motion, which is ascribed to friction among the molecules of the body. This internal friction is called *viscosity*.

The forces to which friction gives rise do not conform to the conditions of conservative forces. They are not uniquely dependent on the position of the moving body, and are exerted only when the body is in motion, and always in such a sense as to oppose the motion. The work done on a body in moving it against friction does not give the body potential energy, and the sum of the kinetic and potential energies in a system, the parts of which exert friction on one another, continually diminishes. Most of the departures from the law of the conservation of mechanical energy exhibited in the ordinary operations of Nature are due to friction. The mechanical energy lost is for the most part transformed into heat.

75. Laws of Friction.—Owing to our ignorance of the arrangement and behavior of molecules, we cannot form a theory of friction based upon mechanical principles. The laws which have been found are almost entirely experimental, and are only approximately true even in the cases in which they apply.

It was found by Coulomb that, when one solid slides over another, the resistance to the motion is proportional to the pressure normal to the surfaces of contact, and is independent of the area of the surfaces and of the velocity with which the moving body slides over the other. It depends upon the nature of the bodies and the character of the surfaces of contact. The ratio of the force required to keep the moving body in uniform motion to the force acting upon it normal to the surfaces of contact is called the *coefficient of friction.*

It was shown experimentally by Poiseuille that the rate of outflow of a liquid from a vessel through a long straight tube of very small diameter is proportional directly to the difference in pressure in the liquid at the two ends of the tube, to the fourth power of the radius of the tube, and inversely to the length of the tube. The flow of liquid under such conditions can be determined by mathematical analysis, and it is found that the results obtained by Poiseuille can only occur if the coefficient of friction between the liquid and the wall of the tube be very great. In other words, we may think of the liquid particles nearest the wall as adhering to it and forming a tube of molecules of the same sort as those of the liquid. The outflow then depends only upon the coefficient of viscosity of the liquid.

The frictional resistance experienced by a solid moving through a liquid or a gas is a function of its velocity. When the motion is slow, approximate results are reached by setting it proportional to the velocity. For higher velocities it is more nearly proportional to their squares, and for very high velocities to still higher powers.

It results from this that the motion of a body falling toward the earth will be resisted by a force that increases as its velocity increases, so that after it has attained a certain velocity the frictional resistance and its weight may become equal, and the body, from that time on, will move with a constant velocity. This explains why rain-drops or falling shot reach the earth with much lower velocities than they would have if there were no friction. Further, since the friction depends on the surface, while the weight is pro-

portional to the volume, the limiting or constant velocity reached is less for small than for large bodies. This explains why the fine drops in a fog or cloud fall so slowly that their motion is scarcely noticed, and why shot return to the ground with small velocities, while the velocity of a returning rifle-ball is still considerable.

From considerations based upon the kinetic theory of gases, Maxwell predicted that the coefficient of viscosity in a gas would be independent of its density. This prediction has been verified by experiment through a wide range of densities. For very low densities it has been shown that this law no longer holds true.

76. Theory of Friction.—The friction between solids is due largely, if their surfaces be rough, to the interlocking of projecting parts. In order to slide the bodies over one another, these projections must either be broken off, or the surfaces must separate until they are released. There is also a direct interaction of the molecules which lie in the surfaces of contact. This appears in the friction of smooth solids, and is the sole cause of the viscosity of liquids and gases. That this molecular action is of importance in producing the friction of solids is seen in the facts that the friction of solids of the same kind is greater than that of solids of different kinds, and that it requires a greater force to start one body sliding over another than to maintain it in motion after it is once started.

CAPILLARITY.

77. Fundamental Facts.—If we immerse one end of a fine glass tube having a very small, or capillary, bore in water, we observe that the water rises in the tube above its general level. We also observe that it rises around the outside of the tube, so that its surface in the immediate vicinity of the tube is curved. If we immerse the same tube in mercury, the surface of the mercury within and just outside the tube, instead of being elevated, is depressed. If we change the tube for one of smaller bore, the water rises higher and the mercury sinks lower within it; but the rise or depression outside the tube remains the same. If we immerse the same tube in different liquids, we find that the heights to which

they ascend vary for the different liquids. If, instead of changing the diameter, we change the thickness of the wall of the tube, no variation occurs in the amount of elevation or depression; and, finally, the rise or depression in the tube varies for any one liquid with its temperature.

78. Law of Force Assumed.—It is found that a force such as is given by the law of mass attraction is not sufficient to produce these phenomena. They can, however, be explained if we assume an additional attraction between the molecules, as we have already done. The expression, then, of the stress between two molecules m and m', at distance r, becomes $F = \frac{mm'}{r^2} + mm'f(r)$.

The only law which it is necessary to assign to the function of r in the second term is, that it is very great at insensible distances, diminishes rapidly as r increases, and vanishes while r, though measurable, is still a very small quantity. For adjacent molecules this molecular attraction is so much greater than the mass attraction, that it is customary, in the discussion of capillary phenomena, to omit the term $\frac{mm'}{r^2}$ from the expression for the force. The distance through which this attraction is appreciable is often called the *range of molecular action*, and is denoted by the symbol ϵ. It is a very small distance, but is assumed to be much greater than the distance between adjacent molecules. Other facts, however, connected with the behavior of gases, lead us to think that the distance between molecules and the range of molecular action are of the same order of magnitude. The theory has not been developed from this point of view, but it is easy to see that the auxiliary idea of surface tension is not incompatible with it, though the precise connection between it and the molecular forces will not have the same form as that given by the older theory.

79. Methods of Development.—The different methods which have been employed to deduce, from this assumed attraction, results which could be submitted to experimental verification, are worthy of notice. They are distinct, though compatible with one

another. Young was the first to treat the subject satisfactorily, though others had given partial and imperfect demonstrations before him. He showed that a liquid can be dealt with as if it were covered at the bounding surface with a stretched membrane, in which is a constant tension tending to contract it. From this basis he proceeded to deduce some of the most important of the experimental laws. Laplace, proceeding directly from the law of the attraction which we have already given, considered the attraction of a mass of liquid on a filament of the liquid terminating at the surface, and obtained an expression for the pressure within the mass at the interior end of the filament. He also was able, not only to account for already observed laws, but to predict, in at least one instance, a subsequently verified result. Some years later, Gauss, dissatisfied with Laplace's assumption, without *a priori* demonstration, of a known experimental fact, treated the subject from the basis of the principle of virtual velocities, which in this case is the equivalent of that of the conservation of energy. He proved that, if any change be made in the form of a liquid mass, the work done or the energy recovered is proportional to the change of surface, and hence deduced a proof of the fact which Laplace assumed, and also an expression for the pressure within the mass of a liquid identical with his. For purposes of elementary treatment the earliest method is still the best. We shall accordingly employ the idea of *surface tension*, after having shown that it may be obtained from the hypothesis of molecular attraction.

80. Surface Tension.—Consider any liquid bounded by a plane surface, of which the line mn (Fig. 28) is the trace, and let the line $m'n'$ be the trace of a parallel plane at the distance ϵ from the plane of mn. Beneath the plane $m'n'$ the liquid will be homogeneous at all points, and the attraction on any one molecule of it due to the surrounding molecules will be the same in all directions. If we consider the

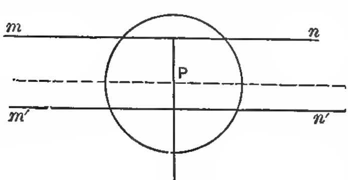

Fig. 28.

attraction for each other of the molecules lying on either side of a small area taken within the body of the liquid, it is easy to see that, if we suppose the molecules on one side of this surface to be removed, while those on the other side are still acted on by the same forces as those which they experienced before, equilibrium may be maintained by the application of a pressure to all points of the surface. On account of the homogeneity of the liquid this pressure will be the same in all directions; when measured for unit of surface it is called the *molecular pressure* in the liquid, and is denoted by K.

The condition of things is different in the surface layer included between the planes $m'n'$ and mn. In the first place, the conditions within the range of molecular action around any one molecule, such as P, are not the same in all directions, and P is therefore acted on by a force which is always normal to the surface, and which is greater when P is nearer the surface mn. This appears at once from the figure, which shows that P is drawn toward the body of the liquid by the molecules contained within the lower hemisphere determined by the range of molecular action, and is drawn upward by those contained within that portion of the upper hemisphere determined by the surface mn and the parallel plane passing through P. In the second place, the pressure on a surface containing P will not be the same for all positions of the surface. It diminishes as P approaches mn, and vanishes at the surface of the liquid. If we consider a surface perpendicular to mn, and suppose the molecules on one side of it removed, it is evident that the forces which act on P will not be normal to the surface and will tend to displace the molecule at P and draw it into the body of the liquid. Such forces give rise to a so-called *tension* in the surface, which tends to contract it. This tension is best conceived of by considering the surface of the liquid interrupted by a thin rigid rod and the liquid removed from one side of it; a force must then be applied to the rod directed away from the liquid, in order to maintain equilibrium. The ratio of the total force applied to

the length of the rod, or the force applied per unit of length, measures the *surface tension.*

81. Energy and Surface Tension.—If the shape of the liquid mass be changed in such a way that its surface increases, work must be done upon those molecules which pass from the interior into the surface. This may either be viewed as work done upon each molecule as it is forced out of the interior mass, where the forces upon it are in equilibrium, into the surface layer, in which it is acted on by a force normal to the surface and in which therefore a movement along that normal involves the doing of work; or it may be looked on as work done against the tension acting in the surface. We call the potential energy gained when the surface increases by one unit the *surface energy* per unit of surface; we will show that it is numerically equal to the surface tension per unit of length.

Suppose a thin film of liquid to be stretched on a frame $ABCD$ (Fig. 29), of which the part BCD is solid and fixed, and the part

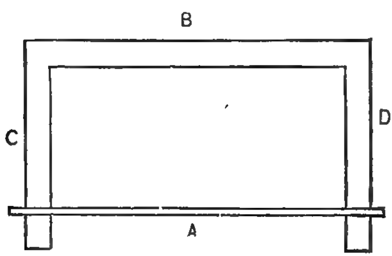

FIG. 29.

A is a light rod, free to slide along C and D. This film tends, as we have said, to diminish its free surface. As it contracts, it draws A towards B. If the length of A be a, and A be drawn towards B over b units, and if E represent the surface energy per unit of surface, the energy lost, or the work done, is expressed by Eab. If we consider the tension acting normal to A, the value of which is T for every unit of length, we have again for the work done during the movement of A, Tab. From these expressions we obtain at once

$E = T$; that is, the numerical value of the surface energy per unit of surface is equal to that of the tension in the surface, normal to any line in it, per unit of length of that line.

82. Equation of Capillarity.—The surface tension introduces modifications in the pressure within the liquid mass (§§ 112 *seq.*) depending upon the curvature of the surface. Consider any infinitesimal rectangle (Fig. 30) on the surface. Let the length of its sides be represented by s and s' respectively, and the radii of curvature of those sides by R and R'. Also let ϕ and ϕ' represent the angles in circular measure subtended by the sides from their respective centres of curvature. Now, a tension T for every unit of length acts normal to s and tangent to the surface. The total tension across s is then Ts; and if this tension be resolved

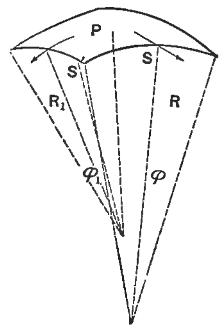

FIG. 30.

parallel and normal to the normal at the point P, the centre of the rectangle, we obtain for the parallel component $Ts \sin \frac{\phi'}{2}$, or, since ϕ' is a very small angle, $Ts\frac{\phi'}{2}$ or $Ts\frac{s'}{2R'}$. The opposite side gives a similar component; the side s' and the side opposite it give each a component $Ts'\frac{s}{2R}$. The total force along the normal at P is then $Tss'\left(\frac{1}{R'} + \frac{1}{R}\right)$; and since ss' is the area of the infinitesimal rectangle, the force or pressure normal to the surface at P referred to unit of surface is $T\left(\frac{1}{R'} + \frac{1}{R}\right)$. From a theorem given by Euler we know that the sum $\frac{1}{R'} + \frac{1}{R}$ is constant at any point for any position of the rectangular normal plane sections; hence the expression we have obtained fully represents the pressure at P.

If the surface be convex, the radii of curvature are positive, and the pressure is directed towards the liquid; if concave, they are

negative, and the pressure is directed outwards. This pressure is to be added to the constant molecular pressure which we have already seen exists everywhere in the mass. If we denote this constant molecular pressure by K, the expression for the total pressure within the mass is $K + T\left(\frac{1}{R'} + \frac{1}{R}\right)$, where the convention with regard to the signs of R' and R must be understood. For a plane surface, the radii of curvature are infinite, and the pressure under such a surface reduces to K.

This equation is known as Laplace's equation.

83. Angles of Contact.—Many of the capillary phenomena appear when different liquids, or liquids and solids, are brought in contact with one another. It becomes, therefore, necessary to know the relations of the surface tensions and the angles of contact. They are determined by the following considerations:

Consider first the case when three liquids meet along a line. Let O represent the point where this line cuts a plane drawn at right angles to it (Fig. 31). Then the tension T_{ab} of the surface of separation of the liquid a from the liquid b, acting normal to this line, is counterbalanced by the tensions T_{ac} and T_{bc} of the surfaces of separation of a and c, b and c. These tensions are always the same for the three liquids under similar conditions of temperature and purity. Knowing the value of the tensions, the angles which they make with one another are determined at once by the parallelogram of forces; and these angles are always constant.

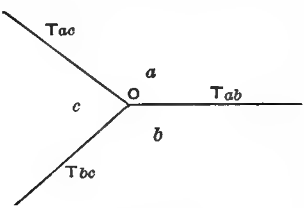

FIG. 31.

Similar relations arise if one of the liquids be replaced by a gas. Indeed, some experiments by Bosscha indicate that capillary phenomena occur at surfaces of separation between gases. We need, therefore, in the subsequent discussions, make no distinction between gases and liquids, and may use the general term fluids.

If T_{ab} be greater than the sum of T_{ac} and T_{bc}, the angle between T_{ac} and T_{bc} becomes zero, and the fluid c spreads itself out in a thin sheet between a and b. Thus, if a drop of oil be placed on water, the tension of the surface of separation between the air and water is greater than the sum of the tensions of the surfaces between the air and oil, and between the oil and water; hence the drop of oil spreads out over the water until it becomes almost indefinitely thin.

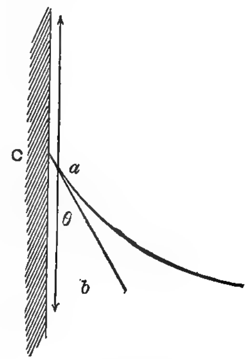

FIG. 32.

In the case of two fluids in contact with a plane solid (Fig. 32), it is evident that the system is in equilibrium when the surface of separation between the fluids a and b makes the angle θ with the solid C given by $T_{ac} = T_{bc} + T_{ab} \cos\theta$. The angle of contact is then determined by the equation $\cos\theta = \dfrac{T_{ac} - T_{bc}}{T_{ab}}$.

If T_{ac} be greater than $T_{ab} + T_{bc}$, the equation gives an impossible value for $\cos\theta$. In this case the angle becomes evanescent, the fluid b spreads itself out, and wets the whole surface of the solid. In other cases the value of θ is finite and constant for the same substances. Thus, a drop of water placed on a horizontal glass plate will spread itself over the whole plate; while a small quantity of mercury placed on the same plate will gather together into a drop, the edges of which make a constant angle with the surface.

84. Plateau's Experiments.—The preceding principles will enable us to explain a few of the most important experimental facts of capillarity.

A series of interesting results was obtained by Plateau from the examination of the behavior of a mass of liquid removed from the action of gravity. His method of procedure was to place a mass of oil in a mixture of alcohol and water, carefully mixed so as to have the same specific gravity as the oil. The oil then had no tendency

to move as a mass, and was free to arrange itself entirely under the action of the molecular forces. Referring to the equation of Laplace, already obtained, it is evident that equilibrium can exist only when the sum $\left(\frac{1}{R'} + \frac{1}{R}\right)$ is constant for every point on the surface. This is manifestly a property of the sphere, and is true of no other finite surface. Plateau found, accordingly, that the freely floating mass at once assumed a spherical form. If a solid body—for instance, a wire frame—be introduced into the mass of oil, of such a size as to reach the surface, the oil clings to it, and there is a break in the continuity of the surface at the points of contact. Each of the portions of the surface divided from the others by the solid then takes a form which fulfils the condition already laid down, that $\left(\frac{1}{R'} + \frac{1}{R}\right)$ equals a constant. Plateau immersed a wire ring in the mass of oil. So long as the ring nowhere reached the surface, the mass remained spherical. On withdrawing a portion of the oil with a syringe, that which was left took the form of two equal *calottes*, or sections of spheres, forming a double convex lens. A mass of oil, filling a short, wide tube, projected from it at either end in a similar section of a sphere. As the oil was removed, the two end surfaces became less curved, then plane, and finally concave.

Plateau also obtained portions of other figures which fulfil the required condition. For example, a mass of oil was made to surround two rings placed at a short distance from one another. Portions of the oil were then gradually withdrawn, when two spherical *calottes* formed, one at each ring, and the mass between the rings became a right cylinder. It is evident that the cylinder fulfils the required condition for every point on its surface.

Plateau also studied the behavior of films. He devised a mixture of soap and glycerine, which formed very tough and durable films; and he experimented with them in air. Such films are so light that the action of gravity on them may be neglected in comparison with that of the surface tension. If the parts of the frame

upon which the film is stretched be all in one plane, the film will manifestly lie in that plane. When, however, the frame is constructed so that its parts mark the edges of any geometrical volume, the films which are taken up by it often meet. Any three films thus meeting arrange themselves so as to make angles of 120° with one another. This follows as a consequence of the proposition which has already been given to determine the equilibrium of surfaces of separation meeting along a line. If four or more films meet, they always meet at a point.

Plateau also measured the pressure of air in a soap-bubble, and found that it differed from the external pressure by an amount which varied inversely as the radius of the bubble. This follows at once from Laplace's equation. This measurement also gives us a means of determining the surface tension; for, from Laplace's equation, the pressure inwards, due to the outer surface, is $T\frac{2}{R}$, and the pressure in the same direction due to the inner surface is also $T\frac{2}{R}$, for the film is so thin that we may neglect the difference in the radii of the two surfaces: hence the total pressure inwards is $\frac{4T}{R}$; and if this be measured by a manometer, we can obtain the value of T.

85. Liquids influenced by Gravity.—Passing now to consider liquid masses acted on by gravity, we shall treat only a few of the most important cases.

If a glass tube having a narrow bore be immersed perpendicularly in water, the water rises in the tube to a height inversely proportional to the diameter of the tube. This law is known as *Jurin's law.*

Let Fig. 33 represent the section of a tube of radius r immersed in a liquid, the surface of which makes an angle θ with the wall. Then if T be the surface tension of the liquid, the tension acting upward is the component of this surface tension parallel to the wall, exerted all around the circumference of the tube. This is

expressed by $2\pi r T\cos\theta$. This force, for each unit area of the tube, is $\frac{2\pi r T\cos\theta}{\pi r^2}$. The downward force, at the level of the free surface, making equilibrium with this, is due to the weight

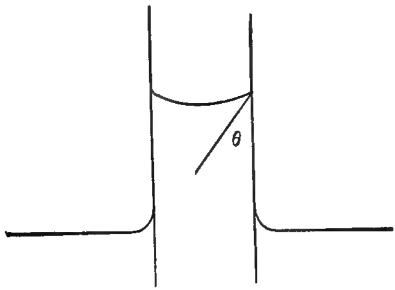

FIG. 33.

of the liquid column (§ 113). If we neglect the weight of the meniscus, this force per unit area, or the pressure, is expressed by hdg, where h is the height of the column and d the density of the liquid. We have, accordingly, since the column is in equilibrium, $\frac{2\pi r}{\pi r^2}T\cos\theta = hdg$; whence $h = \frac{2T\cos\theta}{rdg}$, and the height is inversely as the radius of the tube.

If the liquid rise between two parallel plates of length l, separated by a distance r, the upward force per unit area is given by the expression $\frac{2l}{lr}T\cos\theta$, and the downward pressure by hdg; whence $h = \frac{2T\cos\theta}{rdg}$, and the height to which the liquid will rise between two such plates is equal to that to which it will rise in a tube the radius of which is equal to the distance between the plates.

If the two plates be inclined to one another so as to touch along one vertical edge, the elevated surface takes the form of a rectangular hyperbola; for, let the line of contact of the plates be taken as the axis of ordinates, and a line drawn in the plane of the free surface of the liquid as the axis of abscissas, the elevation correspond-

ing to each abscissa is inversely as the distance between the plates at that point, and the elevations are therefore inversely as the abscissas: hence the product of any abscissa by its corresponding ordinate is a constant. The extremities of the ordinates then mark out a rectangular hyperbola referred to its asymptotes.

86. Movements of Solids.—In certain cases the action of the capillary forces produces movements in solid bodies partially immersed in a liquid. For example, if two plates, which are both either wetted or not wetted by the liquid, be partially immersed vertically, and brought so near together that the rise or depression of the liquid due to the capillary action begins, then the plates will move towards one another. In either case this movement is explained by the inequality of pressure on the two sides of each plate. When the liquid rises between the plates, the pressure is zero at that point in the column which lies in the same plane as the free external surface. At every internal point above this the molecules of the liquid are in a state of negative pressure or tension, and the plates are consequently drawn together. When the liquid is depressed between the plates, they are pressed together by the external liquid above the plane in which the top of the column between the plates lies. When one of the plates is wetted by the liquid and the other not, the plates move apart. This is explained by noting, that, if the plates be brought near together, the convex surface at the one will meet the concave surface at the other, and there will be a consequent diminution in both the elevation and the depression at the inner surfaces of the plates. The elevation and depression at the outer surfaces remaining unchanged, there will result a pull outwards on the wetted plate and a pressure outwards on the plate which is not wetted; and they will consequently move apart. Laplace showed, however, as the result of an extended discussion, that, though seeming repulsion exists between two plates such as we have just considered, yet, if the distance between the plates be diminished beyond a certain value, this repulsion changes to an attraction. This prediction has been completely verified by the most careful experiments.

87. Porous Bodies.—*Porous bodies* may be considered as assemblages of more or less irregular capillary tubes. Thus the explanation of many natural phenomena—as the wetting of a sponge, the rise of the oil in the wick of a lamp—follows directly from the preceding discussion.

DIFFUSION.

88. Solution and Absorption.—Many solid bodies, immersed in a liquid, after awhile disappear as solids, and are taken up by the liquid. This process is called *solution*. The quantity of any body which a unit quantity of a given liquid will dissolve at a given temperature, is called its *solubility* in that liquid at that temperature. The solubility of a given solid varies greatly for different liquids, in many cases being so small as to be inappreciable.

One liquid may also be dissolved in another, the degree of solubility differing very much for different liquids. At ordinary temperatures many liquids are practically insoluble in others, but there is reason to believe that as the liquids approach their critical points (§ 223), their solubilities in other liquids increase, and that at their critical points any liquid is soluble in all others in any proportion.

Gases are also taken into solution by liquids. The process is usually called *absorption*. The quantity of gas dissolved in any liquid depends upon the temperature, and varies directly with the pressure. The solubility of any gas at a given temperature and at standard pressure is called its *coefficient of absorption* at that temperature.

Gases, in general, adhere strongly to the surfaces of solids with which they are in contact. This adhesion is so great, that the gases are sometimes condensed so as to form a dense layer which probably penetrates to some depth below the surface of the solid. The process is called the *absorption* of gases by solids. When the solid is porous, its exposed surface is greatly extended, and hence much larger quantities of gas are condensed on it than would otherwise be the case. When this condensation occurs there is in gen-

eral a rise of temperature, which may be so great as to raise the solid to incandescence. Thus, for example, spongy platinum, placed in a mixture of oxygen and hydrogen, becomes so heated as to inflame it.

89. Free Diffusion of Liquids.—When two liquids which are miscible are so brought together in a common vessel that the heavier is at the bottom and the lighter rests upon it in a well-defined layer, it is found that after a time, even though no agitation occurs, they become uniformly mixed. Molecules of the heavier liquid make their way upwards through the lighter; while those of the lighter make their way downwards through the heavier, in apparent opposition to gravitation. *Diffusion* is the name which is employed to designate this phenomenon and others of a similar nature.

When one of the liquids is colored,—as, for example, solution of cupric sulphate,—while the other is colorless, the progress of the experiment may easily be watched and noted. When both liquids are colorless, small glass spheres, adjusted and sealed so as to have different but determinate specific gravities between those of the liquids employed, may be placed in the vessel used in the experiment, and will show by their positions the degree of diffusion which has occurred at any given time.

90. Coefficient of Diffusion.—Experiment shows that the amount of a salt in solution which at a given temperature passes, in unit time, through unit area of a horizontal surface, depends upon the nature of the salt and the rate of change of concentration at that surface,—that is, the quantity of a salt that passes a given horizontal plane in unit time is κCA, where A is the area, C the rate of change of concentration, and κ a coefficient that depends upon the nature of the substance. By *rate of change of concentration* is meant the difference in the quantities of salt in solution, measured in grams per cubic centimetre, at two horizontal planes one centimetre apart, supposing the concentration to diminish uniformly from one to the other. It is plain, that, if C and A in the above expression be each equal to unity, the quantity of salt passing in

unit time is κ. The quantity κ, called the *coefficient of diffusion*, is, therefore, the quantity of salt that passes in unit time through unit area of a horizontal plane when the difference of concentration is unity. Coefficients of diffusion increase with the temperature, and are found not to be entirely independent of the degree of concentration.

As implied above, the units of mass and length employed in these measurements are respectively the gram and the centimetre; but, since in most cases the quantity of salt that diffuses in one second is extremely small, it is usual to employ the day as the unit time.

91. Diffusion through Porous Bodies.—It was found by Graham that diffusion takes place through porous solids, such as unglazed earthenware or plaster, almost as though the liquids were in direct contact, and that a very considerable difference of pressure can be established between the two faces of the porous body while the rate of diffusion remains nearly constant.

92. Diffusion through Membranes.—If the membrane through which diffusion occurs be of a type represented by animal or vegetable tissue, the resulting phenomena, though in some respects similar, are subject to quite different laws. Colloid substances pass through the membrane very slowly, while crystalloid substances pass more freely. It is to be noted that the membrane is not a mere passive medium, as is the case with the porous substances already considered, but takes an active part in the process; and consequently one of the liquids frequently passes into the other more rapidly than would be the case if the surfaces of the liquids were directly in contact.

If the membrane separate two crystalloids, it often happens that both substances pass through, but at different rates. In accordance with the usage of Dutrochet, we may say there is *endosmose* of the liquid which passes more rapidly to the other liquid, and *exosmose* of the latter to the former. The whole process is frequently called *osmosis*. If the membrane be stretched over the end of a tube, into which the more rapid current sets, and the tube be

placed in a vertical position, the liquid will rise in the tube until a very considerable pressure is attained. Dutrochet called such an instrument an *endosmometer*.

Graham made use of a similar instrument, which he called an *osmometer*, by means of which he studied, not only the action of porous substances, such as are mentioned above, but also that of various organic tissues; and he was able to reach quantitative results of great value. Pfeffer has more recently made an extended study of the phenomena of osmosis, especially in those aspects relating to physiological phenomena. He has shown that colloid membranes produced by purely chemical means are even more efficient than the organic membranes employed by Graham.

93. Dialysis.—Upon the principles just set forth Graham has founded a method of separating crystalloids from any colloid matters in which they may be contained, which is often of great importance in chemical investigations. The apparatus employed by Graham consists of a hoop, over one side of which parchment paper is stretched so as to constitute a shallow basin. In this basin is placed the mixture under investigation, and the basin is then floated upon pure water contained in an outer vessel. If crystalloids be present, they will in due time pass through the membrane into the water, leaving the colloids behind. The process is often employed in toxicology for separating poisons from ingesta or other matters suspected of containing them. It is called *dialysis*, and the substances that pass through are said to *dialyse*.

94. Osmotic Pressure.—Pfeffer carried out a series of investigations to determine whether osmosis produces inequalities of pressure on the opposite sides of the membrane. In his examination of this question Pfeffer used membranes formed by chemical action in the pores of earthenware cells. These cells, which were designed to hold the solution to be examined, could be tightly closed and connected with a manometer, or instrument for measuring pressure. In the typical cases examined, the membranes permitted the solvent to pass through them freely, but were impervious to the

substances in solution. They may be called semi-permeable membranes.

Pfeffer's results may be described most easily by using as an example the case of solutions of cane-sugar and water. When the cell was filled with water containing sugar in solution, and immersed in pure water, the water began to enter the cell from without; the pressure in the cell, as indicated by the manometer, began at once to increase, and continued increasing for some time, until a rather large definite increase of pressure had been reached. This increase of pressure in the cell is called the *osmotic pressure* or *solution pressure* of the solution. The most important laws established by Pfeffer, de Vries, and others concerning the relations of osmotic pressure to the character of the solution and its circumstances are as follows, it being understood that the statements to be made refer to dilute solutions and to solutions which are not electrolytes (§ 279). Solutions which are not electrolytes may be called *indifferent solutions* (§ 285). For solutions which are electrolytes the statements need some modifications.

The osmotic pressure is independent of the nature of the solvent and of the character of the membrane, provided it is impervious to the substance dissolved.

The osmotic pressure is proportional to the concentration of the solution or to the quantity of the dissolved substance contained in unit volume. It increases as the temperature rises, and the relation between the increase of pressure and the rise of temperature is the same as that which obtains for gases (§ 211).

Weights of different substances which are proportional to the molecular weights of those substances contain equal numbers of molecules. Solutions formed by dissolving, in equal quantities of the solvent, masses of different substances proportional to the molecular weights of the substances, therefore contain equal numbers of molecules of these substances. They may be called *equimolecular solutions*. It is found that the osmotic pressure exerted by equimolecular solutions of different substances is the same. Solutions which exert equal osmotic pressures are called *isotonic*.

Solutions which are isotonic at one temperature are isotonic at all temperatures, or, the change of osmotic pressure with temperature is the same for all equimolecular solutions. The absolute value of the osmotic pressure is the same as the pressure which would be exerted by a gas contained in the same vessel as the solution, and having the same number of molecules as there are molecules of the substance dissolved.

It should be stated that these laws have not all been proved conclusively by experiment, but they are well established on theoretical grounds.

95. Dissociation.—The foregoing laws of osmotic pressure do not hold for all solutions. The deviations which appear in solutions which are not highly dilute are explained in the same way as the departure of highly-compressed gases from the similar laws of gaseous pressure, namely, by the absence of those simple conditions upon which only these laws are theoretically possible. The deviations of electrolytes from these laws, which are sometimes very great, have been explained by Arrhenius as the result of the separation of some or all of the molecules of the dissolved substance into their constituent portions or ions. This separation is called *dissociation*. The dissociation theory receives abundant support from the phenomena of electrolysis, and will be discussed in that connection (§ 285).

96. Laws of Diffusion of Gases.—Gases obey the same elementary laws of diffusion as liquids. The rate of diffusion varies inversely as the pressure, directly as the square of the absolute temperature, and inversely as the square root of the density of the gas. A gas diffuses through porous solids according to the same laws. An apparatus by which this may be conveniently illustrated consists of a porous cell, the open end of which is closed by a stopper, through which passes a long tube. This is placed in a vertical position, with the open end of the tube in a vessel of water. If, now, a bell-jar containing hydrogen be placed over the porous cell, hydrogen passes into the cell more rapidly than the air escapes from it: the pressure inside is increased, as is shown by

the escape of bubbles from the end of the tube. If, now, the jar be removed, diffusion outward occurs more rapidly than diffusion inward: the pressure within soon becomes less than the atmospheric pressure, as is shown by the rise of the water in the tube.

The laws of gaseous diffusion have been shown by Osborne Reynolds to be consistent with the kinetic theory of gases.

ELASTICITY.

97. Strain and Stress.—In the discussion of the third law of motion (§ 26) stress was defined as the mutual action of two bodies. In the applications made of the third law up to this point the stress has been considered entirely with reference to the two bodies between which it acts; that is, it has been tacitly assumed that the action is immediate, or, as it is called, is an action at a distance. But in many cases the action between two bodies is manifestly not of this sort, but is due to the presence and action of intervening bodies. These intervening bodies, when looked at generally, are called the intervening medium. In these cases we may apply the third law of motion to the parts of the medium, and assert that there exists a *stress* between any two contiguous portions of the medium. This stress will vary from point to point and with the direction of the surface across which it acts, and also with the peculiarities of the medium. Experiment shows that the application of stress to a medium is always accompanied by a change of form or *deformation* of the medium. This deformation is called a *strain.*

In some bodies equal stresses applied in any direction produce equal and similar strains. Such bodies are *isotropic.* In others the strain alters with tho direction of the stress. These bodies are *eolotropic.*

According to the molecular theory of matter, the form of a body is permanent so long as the resultant of the stresses acting on it from without, with the interior forces existing between the individual molecules of the body, reduces to zero. The molecular forces and motions are such that there is a certain form of the body

for every external stress in which its molecules are in equilibrium. Any change of the stress in the body is accompanied by a readjustment of the molecules, which is continued until equilibrium is again established.

98. Strains.—The complete geometrical representation of the changes of form which occur when a body is strained is in general impossible, or at least exceedingly complicated. In the theory of elasticity it is generally possible to avail ourselves of a simplification in the character of the strain, which facilitates its geometrical representation, by assuming that the strain is such that a line in the body which was straight in its unstrained position remains straight after the strain: such a strain is called a *homogeneous strain.* It may be shown, by an argument too extended for presentation here, that in any case of homogeneous strain there are always three directions in the strained body, at right angles to one another, in which the only change produced by the strain is a change in length and not a change in relative direction. Thus, if the strained body be originally a cube, with its sides parallel to these three directions, the cube will strain into a rectangular parallelepiped. If the strained body be originally a sphere, it will strain into an ellipsoid, the three axes of the ellipsoid being the three directions already mentioned. These three directions are called the *principal axes of strain.*

The increase in length of a line of unit length by strain is called its *elongation.* Evidently, from the description of the relations of a homogeneous strain to the principal axes, the whole strain will be described if the elongations along the principal axes be given. Let us denote by e_1, e_2, e_3 the elongations, which may be either positive or negative, along the three principal axes. These elongations are assumed to be so small in comparison with the unit line that their squares or products may be neglected. Then, in the examples just given, if a represent a side of the cube before strain and a^3 its volume, the increase in volume of the cube by the strain is given by $a^3(1 + e_1)(1 + e_2)(1 + e_3) - a^3 = a^3(e_1 + e_2 + e_3)$, since the products of the e's may be neglected. Similarly, the sphere,

of which the radius is r, becomes by the strain the ellipsoid, of which the axes are $r(1+e_1)$, $r(1+e_2)$, $r(1+e_3)$; the increase in volume of the sphere by the strain is therefore

$$\tfrac{4}{3}\pi r^3(1+e_1)(1+e_2)\,(1+e_3)-\tfrac{4}{3}\pi r^3=\tfrac{4}{3}\pi r^3(e_1+e_2+e_3).$$

The quantity $e_1+e_2+e_3$ is called the *coefficient of expansion* of the body.

Two cases of strain need to be specially examined—the pure expansion or *dilatation*, and the *shear* or shearing strain. A dilatation occurs if the three coefficients of elongation are equal; in this case the strained cube remains a cube, the strained sphere remains a sphere, and the change of volume in each case is $3e$ times the original volume. A shear occurs when one of the coefficients, say e_3, equals zero, and when e_1 equals $-e_2$; in this case the expansion is zero.

The shear may be defined from another point of view. For, consider a body subjected to a shear and suppose a section made in it by the plane containing the elongations e and $-e$: it is clear that the shear will be completely described if we describe the deformation of a figure in this plane. We select for this purpose a rhombus, $ABDC$, of which the diagonals AD and BC are so related that after the shear we have $AD(1+e)=BC$ and $BC(1-e)=AD$. If the rhombus produced by the shear be turned until one of its sides coincides with AB, we shall have the original rhombus and the one produced by shear in the relation shown in Fig. 34. The new rhombus $AC'D'B$ may manifestly be produced from the original rhombus by the displacement of all its lines parallel to the fixed base AB, each line being displaced by an amount proportional to its distance from the line AB. The ratio of this displacement to the distance of the displaced line from the base AB is called the *amount* of the shear; that is, $\dfrac{DD'}{EB}$ is the amount of the shear.

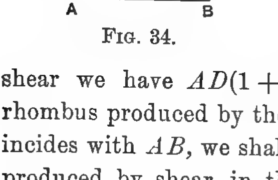

FIG. 34.

99. The Superposition of Strains.—We will now show that two elongations, applied successively or simultaneously in the same direction, are equivalent to a single elongation equal to their sum. This follows from the assumption already made, that the elongations are so small that their squares or products may be neglected. For, suppose a line of unit length to receive the elongation e_1; its length becomes $1 + e_1$. If it then receive the elongation e_2, its length becomes $1 + e_1 + e_2(1+e_1) = 1 + e_1 + e_2$, because the product $e_1 e_2$ may be neglected. This principle is called the principle of the *superposition of strains.*

By its help we may show that a simple elongation may be produced by the combination of a dilatation and two equal shears in planes at right angles to each other. In the case of a simple elongation, the elongations along the principal axes are e, 0, 0. Let us suppose a dilatation of which the elongations are $\frac{e}{3}, \frac{e}{3}, \frac{e}{3}$; a shear of which the elongations are $\frac{e}{3}, -\frac{e}{3}$, 0; and a shear of which the elongations are $\frac{e}{3}$, 0, $-\frac{e}{3}$. By the principle of the superposition of strains we find the elongations produced if these three strains be superposed by adding the three elongations along the three axes. Carrying out this operation we obtain e, 0, 0 as the elongations produced by the superposition, that is, the superposition of these three strains is equivalent to a simple elongation. Since all homogeneous strains may be produced by three simple elongations at right angles to each other, any homogeneous strain may be produced by a combination of dilatations and shears.

100. Stresses.—If a body be maintained in equilibrium by forces applied to points on its surface, and if we conceive it divided into two parts, A and B, by an imaginary surface drawn through it, and if we assume, for the present, the molecular structure of matter, it is clear that the forces applied to the portion A of the body are in equilibrium with the forces which act between the molecules of A lying near the surface which divides it from B, and the mole-

cules of B lying on the other side of that surface. Similarly, the forces which act on B are in equilibrium with the forces which act across the surface between the molecules of B and A. Let us consider any area s taken in the surface separating A and B. Represent by F the sum of the molecular forces which act across that area. If the forces which act across different equal elements of the area be equal, the ratio $\frac{F}{s}$ is called generally the *pressure per unit area* on the surface s, or, simply, the *pressure* on the surface. This pressure is positive if the force F be directed away from the portion of the whole body which is held in equilibrium, negative if directed toward that portion. It is plain, from the equality of action and reaction, that if this force be directed toward the portion A of the body, an equal force is directed toward the portion B at every point of the surface which separates A and B.

The name *pressure* is frequently reserved for a negative pressure in the sense just defined; when the pressure is positive, it is frequently called a *tension*. In case the force which acts across the surface between A and B vary from element to element of that surface, the *pressure at a point* of the surface is the limit of the ratio $\frac{F}{s}$, when the area s is so drawn that its centre of inertia is always kept at that point, and is diminished indefinitely.

The forces acting across the surface separating A and B will, in general, make different angles with the surface at the different points of it. Similarly, the pressure which is substituted for the forces makes different angles with the surface at different points. The pressure, being a vector quantity, like the force from which it is derived, may be resolved into components perpendicular to the surface and in the plane tangent to it. It is best, for the sake of greater generality in our statements, to consider the tangential component of pressure as resolved into two components, at right angles to each other in the tangent plane. These components are called respectively, the *normal pressure* and the *tangential pressures*.

To examine the relations which must hold among the components of pressure in different directions at any point within a body subjected to stress, we consider a small cube described in a body, and examine the relations among the pressures on its faces necessary to maintain it in equilibrium. We assume that no external forces act directly on the matter contained in the cube. In general, each of the faces of the cube will be subjected to a stress. This stress may be resolved into a normal component and two tangential components taken parallel with the sides of the face to which the stress is applied. Calling the normal components acting on two opposite faces P and P', those acting on another pair of opposite faces Q and Q', and those acting on the third pair R and R', we may express the conditions that the centre of mass of the cube will not be displaced by the equations $P=P'$, $Q = Q'$, $R = R'$.

Since the forces which act upon the cube are in equilibrium, and since their normal components maintain the equilibrium of the centre of mass, their tangential components give rise to couples, and these couples are also in equilibrium. These couples are arranged as shown in Fig. 35, for those lying in the plane of one pair of faces. Since equilibrium exists, the two couples formed by the forces S and the forces S' are equal, and therefore $S = S'$, where S and S' may be used to denote the tangential pressures on the surfaces of the cube. Similar couples in equilibrium will act on the cube in two other planes at right angles with this one so that the whole set of pressures acting on the cube are the three normal pressures P, Q, R, and the three tangential pressures S, T, U. It may be shown, by an analytical method that need not be given, that if a small sphere be described about a point in the body and the pressures applied to its surfaces examined, there will be three radii at right angles to each other, at the extremities of which the pressures are normal to the surface of the sphere. These three directions are called the *principal axes of stress.*

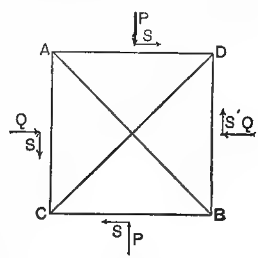

FIG. 35.

The combination of tangential stresses which maintain equilibrium may be considered from another point of view. For, if we examine the triangular prism of which the cross-section is ABD (Fig. 35), and to which the tangential stresses S and S' are applied, it appears at once that equilibrium will obtain when a force equal to the resultant of aS and aS', where a is the area of each of the square faces of the prism, is applied to the face of which AB is the trace. The area of this face is $a\sqrt{2}$, and if X represent the pressure on this face, the force applied to it is $aX\sqrt{2}$. But S equals S', and the resultant of aS and aS' is $aS\sqrt{2}$; whence $X = S$. A similar pressure acts in the opposite direction upon the face of the similar prism ACB. These pressures are positive, that is, they are tensions which tend to separate the parts of the body to which they are applied. If we compound the tangential stresses in another manner by taking as the element of the combination the stresses applied to the faces AD and AC, it is at once evident that they are equivalent to a negative pressure S upon the diagonal face CD. A similar pressure acts across the same face toward the other prism CBD. We may therefore consider the set of stresses constituting the couples in the plane $ACBD$ as equivalent to a positive pressure or tension in the direction of one diagonal and a negative pressure in the direction of the other diagonal. This combination of couples, or its equivalent tension and pressure, is called a *shearing stress.*

101. Superposition of Stresses.—Stresses, whether pressures or tensions, being vector quantities, are compounded like other vector quantities, and, in particular, when they are in the same line, are added algebraically.

Suppose a cube so subjected to stress that equal and opposite pressures, which we will assume to be directed outward from the cube, act on two opposite faces, and that the other faces experience no stress. Such a stress is called a *longitudinal traction.* We will show that this form of stress may be obtained by the combination of a stress made up of equal tensions acting on each face of the cube, and of two shearing stresses.

In Fig. 36 let P represent the value of the longitudinal

traction. It may be considered as made up of three equal tractions $\frac{P}{3}$. Apply to each of the four other faces of the cube two opposite stresses, each equal to $\frac{P}{3}$. Two of these pairs of stresses are represented in the figure. These stresses on the sides of the cube, being equal and opposite, are equivalent to no stress. It is evident that the combination of stresses here described is equivalent to a tension $\frac{P}{3}$ applied to each face of the cube, to a shearing stress $\frac{P}{3}$ acting in the plane of the figure,

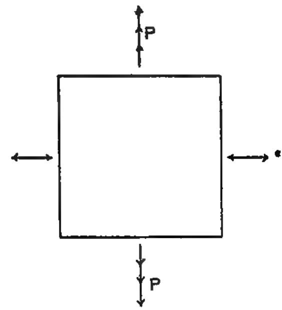

FIG. 36.

and to a shearing stress $\frac{P}{3}$ acting in the plane at right angles to the plane of the figure. Thus the longitudinal traction may be resolved into a tension uniform in all directions and two shearing stresses, all of the same numerical value.

The uniform tension just employed is an example of a hydrostatic stress. More generally, a *hydrostatic stress* is a stress which is normal to any surface element drawn in a body, whatever be its direction. The numerical value of a hydrostatic stress is the same in whatever direction the surface be drawn to which it is applied. To show this, we examine the relations of the pressures on the faces of the tetrahedron formed by passing a plane through the points ABC taken infinitely near the point O (Fig. 37) on lines drawn through that point in the directions of the three coordinate axes. Let l, m, n represent the direction cosines of the normal to the face ABC, and let a represent the area of this face; the areas of the other faces are respectively equal to al, am, an. Let X, P, Q, R represent the pressures on the faces in the order mentioned: the forces acting on the faces are then Xa, Pal, Qam, and Ran. By the definition of hydrostatic

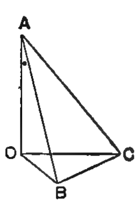

FIG. 37.

stress these forces are normal to their respective faces, and the tetrahedron will be in equilibrium when the components of the force X are equal respectively to the forces applied to the other faces; that is, when $Xa.l = P.al$, $Xa.m = Q.am$, $Xa.n = R.an$; that is, when $X = P = Q = R$.

It has been stated that the stresses in a body may always be represented by the combination of three longitudinal stresses at right angles to each other. Since a longitudinal stress may be replaced by a hydrostatic stress and two shearing stresses, it follows that any stress in a body may be replaced by a hydrostatic stress and a proper combination of shearing stresses.

102. Relations of Stress and Strain. Modulus of Elasticity.— When a body serves as the medium for the transmission of stress it experiences a deformation or strain, the type of strain depending upon the stress applied. The resistance offered by a body to deformation is ascribed to its *elasticity*. If the body be deformed in a definite way by a given stress, and recover its original condition when the stress is removed, it is said to be *perfectly elastic*. If the deformation of a body do not exceed the limits within which it may be considered perfectly elastic, it may be proved by experiment that the strain is of the same type as the stress and proportional to it. This law was proved for certain cases by Hooke, and is known as *Hooke's Law*.

The ratio of the stress applied to the strain experienced by a unit of the body measures the elasticity of the substance composing the body. This ratio is called the *modulus of elasticity* of the body, or simply its *elasticity;* its reciprocal is the *coefficient of elasticity*. It is of course understood that the stress and strain are of the same type. Thus, for example, the voluminal elasticity of a fluid is measured by the ratio of any small change of pressure to the corresponding change of unit volume. The tractional elasticity of a wire stretched by a weight is measured by the ratio of any small change in the stretching weight to the corresponding change in unit length.

Since all stresses may be reduced to hydrostatic stresses and

shearing stresses, and all strains to dilatations and shearing strains, the knowledge of the *voluminal elasticity* and of the elasticity exhibited during a shear, or the *rigidity*, is sufficient to describe the elasticity of the body under any form of stress.

103. Voluminal Elasticity.—Let a body of volume V be subjected to a uniform hydrostatic stress P, by which it undergoes a change of volume, given by v. From Hooke's law we know that, at least within certain limits of stress and consequent deformation, v and P are proportional. The dilatation or the change of the unit of volume is $\frac{v}{V}$. The modulus of elasticity in this case, or the *voluminal elasticity* of the body, is therefore $\frac{PV}{v}$. The voluminal elasticity is denoted by k.

104. Rigidity.—Let S be one of the tangential stresses which constitute a simple shearing stress, that is, a shearing stress of which the elements act in one plane; then the deformation produced is a simple shear. The modulus of *rigidity* is measured by the ratio of the shearing stress to the amount of the shear (§ 98); it is denoted by n.

The amount of the shear may be defined in a more convenient form as follows: Let us suppose that the rhombus $ACDB$ (Fig. 38) has been strained by a simple shear into the rhombus $AC'D'B$, and that this deformation is infinitesimal. The elongation of the diagonal AD is then FD'. The triangle DFD' is then an isosecles triangle, since the angle DFD' is a right angle, and the angle $DD'F$ differs from half a right angle only by an infinitesimal. Therefore $FD'\sqrt{2} = DD'$. Now AD, being the diagonal of a rhombus that is only infinitesimally different from a square, is equal to $BE\sqrt{2}$; and therefore the amount of the shear, or $\frac{DD'}{BE}$, equals $\frac{2FD'}{AD}$, that is, equals twice the elongation along the axis of the shear.

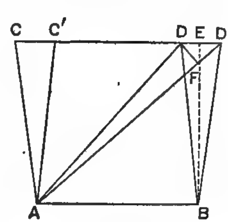

FIG. 38.

The modulus of rigidity is therefore equal to half the tangential stress S divided by the elongation of unit length along the axis of the shear.

105. Modulus of Voluminal Elasticity of Gases.—Within certain limits of temperature and pressure, the volume of any gas, at constant temperature, is inversely as the pressure upon it. This law was discovered by Boyle in 1662, and was afterwards fully proved by Mariotte. It is known, from its discoverer, as *Boyle's law.*

Thus, if p and p' represent different pressures, v and v' the corresponding volumes of any gas at constant temperature, then

$$pv = p'v'. \tag{45}$$

Now, $p'v'$ is a constant which may be determined by choosing any pressure p' and the corresponding volume v' as standards: hence we may say, that, at any given temperature, the product pv is a constant. The limitations to this law will be noticed later.

Let p and v represent the pressure and volume of a unit mass of gas at a constant temperature. A small increase Δp of the pressure will cause a diminution of volume Δv; by Boyle's law we have the relation $pv = (p + \Delta p)(v - \Delta v) = pv + v\Delta p - p\Delta v - \Delta p\Delta v$. We may assume that the increment Δp is very small, in which case Δv will also be small; we may therefore, in the limit, neglect the product of these increments and obtain $\frac{\Delta p}{\Delta v} = \frac{p}{v}$. Now $\frac{\Delta v}{v}$ is the change of unit volume, and therefore $\frac{\Delta p}{\Delta v}v = p$ is the modulus of voluminal elasticity. The elasticity of a gas at constant temperature is therefore equal to its pressure.

106. Modulus of Voluminal Elasticity of Liquids.—When liquids are subjected to voluminal compression, it is found that their modulus of elasticity is much greater than that of gases. For at least a limited range of pressures the modulus of elasticity of any one liquid is constant, the change in volume being proportional to

the change in the pressure. The modulus differs for different liquids.

The instrument used to determine the modulus of elasticity of liquids is called a *piezometer*. The first form in which the instrument was devised by Oersted, while not the best for accurate determinations, may yet serve as a type.

The liquid to be compressed is contained in a thin glass flask, the neck of which is a tube with a capillary bore. The flask is immersed in water contained in a strong glass vessel fitted with a water-tight metal cap, through which moves a piston. By the piston, pressure may be applied to the water, and through it to the flask and to the liquid contained in it.

The end of the neck of the small flask is inserted downwards under the surface of a quantity of mercury which lies at the bottom of the stout vessel. The pressure is registered by means of a compressed-air manometer (§ 124) also inserted in the vessel. When the apparatus is arranged, and the piston depressed, a rise of the mercury in the neck of the flask occurs, which indicates that the water has been compressed.

An error may arise in the use of this form of apparatus from the change in the capacity of the flask, due to the pressure. Oersted assumed, since the pressure on the interior and exterior walls was the same, that no change would occur. Poisson, however, showed that such a change would occur, and gave a formula by which it might be calculated. By introducing the proper corrections, Oersted's piezometer may be used with success.

A different form of the instrument, employed by Regnault, is, however, to be preferred. In it, by an arrangement of stopcocks, it is possible to apply the pressure upon either the interior or exterior wall of the flask separately, or upon both together, and in this way to experimentally determine the correction to be applied for the change in the capacity of the flask.

It is to be noted that the modulus of elasticity for liquids is so great, that, within the ordinary range of pressures, they may be

regarded as incompressible. Thus, for example, the alteration of volume for sea-water by the addition of the pressure of one atmosphere is 0.000044. The change in volume, then, at a depth in the ocean of one kilometre, where the pressure is about 99.3 atmospheres, is 0.00437, or about $\frac{1}{230}$ of the whole volume.

107. Modulus of Voluminal Elasticity of Solids.—The modulus of voluminal elasticity of solids is believed to be generally greater than that of liquids, though no reliable experimental results have yet been obtained.

The modulus, as with liquids, differs for different bodies.

108. Elasticity of Traction.—The first experimental determinations of the relations between the elongation of a solid and a tension acting on it were made by Hooke in 1678. Experimenting with wires of different materials, he found that for small tractions the elongation is proportional to the stress. It was afterwards found that this law is true for small compressions.

The ratio of the stretching weight to the elongation of unit length of a wire of unit section is the *modulus of tractional elasticity.* For different wires it is found that the elongation is proportional to the length of the wires and inversely to their section. The formula embodying these facts is

$$e = \frac{Sl}{\mu s}, \tag{46}$$

where e is the elongation, l the length, s the section of the wire, S the stretching weight, and μ the modulus of tractional elasticity.

The behavior of a body under traction may be examined in the following way: We assume for convenience that the traction is applied to the upper and lower faces of a cube with sides of unit length. As already shown, the traction P is equivalent to a hydrostatic tension $\frac{P}{3}$ and two shearing stresses equivalent to two tensions $\frac{P}{3}$ in the direction of the traction, and a pressure $\frac{P}{3}$ in each of two directions at right angles to this and to each other. The hydro-

static tension causes an increase of volume given by $\frac{P}{3k}$. This is equivalent to an elongation of each side of the cube equal to $\frac{P}{9k}$, since the changes of form are supposed infinitesimal (§ 98). One of the shears produces an elongation between the upper and lower faces equal to $\frac{P}{6n}$, and a negative elongation or contraction equal to $\frac{P}{6n}$ between one pair of the other faces. The other shear produces an equal elongation between the upper and lower faces and an equal contraction between the remaining pair of faces. The total elongation between the upper and lower faces is therefore $P\left(\frac{1}{9k}+\frac{1}{3n}\right)$, and the total contraction between either pair of the other faces is given by $P\left(\frac{1}{6n}-\frac{1}{9k}\right)$.

Since the two shears involved in the longitudinal traction cause no change of volume, the change of volume experienced by the body is due to the hydrostatic tension alone. It is therefore equal to $\frac{P}{3k}$. A body under longitudinal traction will therefore experience an increase of volume unless it is practically incompressible, that is, unless the ratio $\frac{P}{k}$ is negligible.

109. Elasticity of Torsion.—When a cylindrical wire, clamped at one end, is subjected at the other to the action of a couple, the axis of which is the axis of the cylinder, it is found that the *amount of torsion*, measured by the angle of displacement of the arm of the couple, is proportional to the moment of the couple, to the length of the wire, and inversely to the fourth power of its radius. It also depends on the modulus of rigidity. The relation among these magnitudes may be shown to be represented by the formula

$$\tau = \frac{2Cl}{\pi n r^4}, \tag{47}$$

where τ is the amount of torsion, l the length, r the radius of the wire, C the moment of couple, and n the modulus of rigidity. No general formula can be found for wires with sections of variable form.

The laws of torsion in wires were first investigated by Coulomb, who applied them in the construction of an apparatus called the *torsion balance*, of great value for the measurement of small forces.

The apparatus consists essentially of a small cylindrical wire, suspended firmly from the centre of a disk, upon which is cut a graduated circle. By the rotation of this disk any required amount of torsion may be given to the wire. On the other extremity of the wire is fixed, horizontally, a bar, to the ends of which the forces constituting the couple are applied. Arrangements are also made by which the angular deviation of this bar from the point of equilibrium may be determined. When forces are applied to the bar, it may be brought back to its former point of equilibrium by rotation of the upper disk. Let Θ represent the *moment of torsion;* that is, the couple which, acting on an arm of unit length, will give the wire an amount of torsion equal to a radian, C the moment of couple acting on the bar, τ the amount of torsion measured in radians; then $C = \Theta\tau$. We may find the value of Θ in absolute measure by a method of oscillations analogous to that used to determine g with the pendulum.

A body of which the moment of inertia can be determined by calculation is substituted for the bar, and the time T of one of its oscillations about the position of equilibrium observed.

Since the amount of torsion is proportional to the moment of couple, the oscillating body has a simple harmonic motion.

The angular acceleration α of the oscillating body is given by the equation $C = \Theta\tau = I\alpha$ (§ 39). Now, since every point in the body has a simple harmonic motion, in which its displacement is proportional to its acceleration, and since its displacement and acceleration are proportional respectively to the angular displacement τ and the angular acceleration α, we may set $\alpha = \frac{4\pi^2}{T^2}\tau$. Making

this substitution, we obtain $\Theta = \frac{4\pi^2 I}{T^2}$, or

$$\Theta = \frac{\pi^2 I}{t^2}, \tag{48}$$

if we observe the single instead of the double oscillation.

The torsion balance may therefore be used to measure forces in absolute units.

If the value of Θ just obtained be substituted in equation (47), we obtain

$$n = \frac{2\Theta l}{\pi r^4} = \frac{2\pi I l}{r^4 t^2}. \tag{49}$$

Since all these magnitudes may be expressed in absolute units, we may obtain the value of n, the rigidity, by observing the oscillations of a wire of known dimensions, carrying a body of which the moment of inertia is known.

110. Elasticity of Flexure.—If a rectangular bar be clamped by one end, and acted on at the other by a force normal to one of its sides, it will be bent or flexed. The amount of flexure—that is, the amount of displacement of the extremity of the bar from its original position—is found to be proportional to the force, to the cube of the length of the bar, and inversely to its breadth, to the cube of its thickness, and to the modulus of tractional elasticity. The formula expressing the relations of these magnitudes is

$$f = \frac{4Fl^3}{\mu b d^3} \tag{50}$$

111. Limits of Elasticity. — The theoretical deductions and empirical formulas which we have hitherto been considering are strictly applicable only to perfectly elastic bodies. It is found that the voluminal elasticity of fluids is perfect, and that within certain limits of deformation, varying for different bodies, we may consider both the voluminal elasticity and the rigidity of solids to be practically perfect for every kind of strain. If the strain be carried beyond the limits of perfect elasticity, the body is permanently deformed. This permanent deformation is called *set*.

Upon these facts we may base a distinction between solids and fluids : a *solid* requires the stress acting on it to exceed a certain limit before any permanent set occurs, and it makes no difference how long the stress acts, provided it lies within the limit. A *fluid*, on the contrary, may be permanently deformed by the slightest shearing stress, provided time enough be allowed for the movement to take place. The fundamental difference lies in the fact that fluids have no rigidity and offer no resistance to shearing stress other than that due to internal friction or viscosity.

A solid, if it be deformed by a slight stress, is *soft;* if only by a great stress, is hard or *rigid*. A fluid, if deformed quickly by any stress, is *mobile ;* if slowly, is *viscous*.

It must not be understood, however, that the behavior of elastic solids under stress is entirely independent of time. If, for example, a steel wire be stretched by a weight which is nearly, but not quite, sufficient to produce an immediate set, it is found that, after some time has elapsed, the wire acquires a permanent set. If, on the other hand, a weight be put upon the wire somewhat less than is required to break it, by allowing intervals of time to elapse between the successive additions of small weights, the total weight supported by the wire may be raised considerably above the *breaking-weight*. If the weight stretching the wire be removed, the return to its original form is not immediate, but gradual. If the wire carrying the weight be twisted, and the weight set oscillating by the torsion of the wire, it is found that the oscillations die away faster than can be explained by any imperfections in the elasticity of the wire.

These and similar phenomena are manifestly dependent upon peculiarities of molecular arrangement and motion. The last two are exhibitions of the so-called viscosity of solids. The molecules of solids, just as those of liquids, move among themselves, but with a certain amount of frictional resistance. This resistance causes the external work done by the body to be diminished, and the internal work done among the molecules becomes transformed into heat.

CHAPTER IV.

MECHANICS OF FLUIDS.

112. Pascal's Law.—A *perfect fluid* may be defined as a body which offers no resistance to shearing-stress. No actual fluids are perfect. Even those which approximate that condition most nearly, offer resistance to shearing-stress, due to their viscosity. With most, however, a very short time only is needed for this resistance to vanish; and all mobile fluids at rest can be dealt with as if they were perfect, in determining the conditions of equilibrium. If they are in motion, their viscosity becomes a more important factor.

As a consequence of this definition of a perfect fluid follows a most important deduction. In a fluid in equilibrium, not acted on by any outside forces except the pressure of the containing vessel, the pressure at every point and in every direction is the same. This law was first stated by Pascal, and is known as *Pascal's law.*

The truth of Pascal's law appears at once from what has been proved about hydrostatic stress (§ 101). For since the fluid offers no resistance to a shearing stress, the only stress within it on any surface must be perpendicular to that surface, and hence has the same value in all directions at a point. To compare the pressure at any two points we draw a line joining them, and, with it as an axis, describe a right cylinder with an infinitesimal radius, and through the two points take cross-sections normal to the axis. Then the pressures on the cylindrical surface being everywhere normal to it, have no tendency to move it in the direction of its axis, and since it is in equilibrium, the pressures on its end surfaces must be equal.

If a vessel filled with a fluid be fitted with a number of pistons of equal area A, and a force Ap be applied to one of them, acting inwards, a pressure Ap will act outwards upon the face of each of the pistons. These pressures may be balanced by a force applied to each piston. If $n + 1$ be the number of the pistons, the outward pressure on n of them, caused by the force applied to one, is npA.

The fluid will be in equilibrium when a pressure p is acting on unit area of each piston. It is plain that the same reasoning will hold if the area of one of the pistons be A and of another be nA. A pressure Ap on the one will balance a pressure of nAp on the other. This principle governs the action of the *hydrostatic press*.

113. Relations of Fluid Pressures due to Outside Forces.—If forces, such as gravitation, act on the mass of a fluid from without, Pascal's law no longer holds true. For, suppose the fluid to be acted on by gravity, and consider a cylinder of the fluid, the axis of which is vertical, and which is terminated by two normal cross-sections. The pressure on the cylindrical surface, being everywhere normal to it, has no effect in sustaining the weight of the cylinder. The weight is sustained wholly by the pressure on the lower cross-section, and must be equal to the difference between that pressure and the pressure on the upper cross-section. As the height of the cylinder may be made as small as we please, it appears that, in the limit, the pressure on the two cross-sections only differs by an infinitesimal; that is, the pressure in a fluid acted on by outside forces is the same at one point for all directions, but varies continuously for different points.

If, in a fluid acted on by gravity, a surface be considered which is everywhere perpendicular to the lines of gravitational force, the pressure at every point in this surface is the same. To show this we draw a line in the surface between any two points of it, and construct around it as axis a cylinder terminated at the chosen points by end-surfaces drawn normal to the axis. The pressures on the cylindrical surface, being normal to it, occasion resultant

forces which are everywhere in the opposite direction to the gravitational force and make equilibrium with it. The cylinder being in equilibrium, by hypothesis, the forces on the end surfaces, which alone can produce movement in the direction of the axis, must also be equal, and the pressures on those surfaces are therefore equal. Surfaces of equal pressure are equipotential surfaces; in small masses of liquid they are horizontal planes; in larger masses, such as the oceans, they are curved so as to be always at right angles to the divergent lines of force.

The surface of separation between two fluids of different densities in a field in which the lines of gravitational force may be supposed parallel is a horizontal plane. For, take two points, a and c, in the same horizontal plane in the lower fluid, and from them draw equal vertical lines terminated at the points b and d, respectively, in the upper fluid. The horizontal planes containing a and c, b and d, respectively, are surfaces of equal pressure. Now with these lines as axes construct right cylinders with the same small radius and terminated by equal cross-sections in the upper and lower horizontal planes. The pressures on the cylindrical surfaces, being everywhere normal to them, will have no effect in sustaining the weights of these cylinders. Their weights are sustained by the difference in pressure between the upper and lower cross-sections, and, since these cross-sections are in surfaces of equal pressure, the difference of pressure is the same for both cylinders, and the weights of the cylinders are therefore equal. By the construction the cylinders contain portions of both the fluids, and since these fluids are of different densities the weights in the cylinders can only be the same when each cylinder contains the same quantity of each fluid, that is, when the surface of separation between the fluids is parallel with the planes which contain the end cross-sections. The surface of separation is therefore also a horizontal plane. This theorem may be extended so as to prove that the surface of separation between two fluids in any gravitational field is at right angles to the lines of gravitational force, or is an equipotential surface.

In an incompressible fluid or liquid the pressure at any point is proportional to its depth below the surface. For, the weight of a column of the liquid contained in a vertical cylinder, terminated by the free surface and by a horizontal cross-section containing the point, is manifestly proportional to the height of the cylinder; and this weight is sustained by the pressure on the lower end cross-section, which must therefore be proportional to the height of the cylinder.

If the height of the cylinder be h and the area of its cross-section s, and if the density of the liquid be D, the weight of the column is $Dshg$. If p represent the pressure at the base, the upward force on the base is ps; so that we have

$$p = Dhg. \tag{51}$$

From the foregoing principles it is evident that a liquid contained in two communicating vessels of any shape whatever, will stand at the same level in both. If, however, a liquid like mercury be contained in the vessels, and if another liquid, like water, which does not mix with it, be poured into one of the vessels, the surface of separation will sink, and the free surface in the other vessel will rise to a certain point. If a horizontal plane be passed through the surface of separation between the two liquids, the pressures at all points of it within the liquids, in both vessels, will be the same. These pressures, which are due to the superincumbent columns of liquid in the two vessels, are given by Dgh and $D'gh'$, and since they are equal, we have $Dh = D'h'$; that is, the heights of the two columns above the horizontal plane passing through the surface of separation are inversely as the densities of the liquids.

There is nothing in this demonstration which requires us to consider both the columns as liquid: one of them may be of any fluid, and equilibrium will obtain when the pressure exerted by that fluid on the surface of separation is equal to the pressure exerted by the column of liquid in the other vessel on the horizontal plane containing the surface of separation; so that, if we know the

density and the height of the liquid column, the pressure exerted by the fluid may be measured.

114. The Barometer.—The instrument which illustrates these principles, and is also of great importance in many physical investigations, is the *barometer*. It was invented by Torricelli, a pupil of Galileo. The fact that water can be raised in a tube in which a complete or partial vacuum has been made was known to the ancients, and was explained by them, and by the schoolmen after them, by the maxim that "Nature abhors a vacuum." They must have been familiar with the action of pumps, for the force-pump, a far more complicated instrument, was invented by Ctesibius of Alexandria, who lived during the second century B.C. It was not until the time of Galileo, however, that the first recorded observations were made that the column of water in a pump rises only to a height of about 10.5 metres. Galileo failed to give the true explanation of this fact. He had, however, taught that the air has weight; and his pupil Torricelli, using that principle, was more successful.

He showed, that if a glass tube sealed at one end, over 760 millimetres long, were filled with mercury, the open end stopped with the finger, the tube inverted, and the unsealed end plunged beneath a surface of mercury in a basin, on withdrawing the finger the mercury in the tube sank until its top surface was about 760 millimetres above the surface of the mercury in the basin. The specific gravity of the mercury being 13.59, the pressure of the mercury column and that of the water column in the pump agreed so nearly as to show that the maintenance of the columns in both cases was due to a common cause,—the pressure of the atmosphere. This conclusion was subsequently verified and established by Pascal, who requested a friend to observe the height of the mercury column at the bottom and at the top of a mountain. On making the observation, the height of the column at the top was found to be less than at the bottom. Pascal himself afterwards observed a slight though distinct diminution in the height of the column on ascending the tower of St. Jacques de la Boucherie in Paris.

The form of barometer first made by Torricelli is still often used, especially when the instrument is stationary, and is intended to be one of precision. In the finest instruments of this class a tube is used which is three or four centimetres in diameter, so as to avoid the correction for capillarity. A screw of known length, pointed at both ends, is arranged so as to move vertically above the surface of the mercury in the cistern. When an observation is to be made, the screw is moved until its lower point just touches the surface. The distance between its upper point and the top of the column is measured by means of a cathetometer; and this distance, added to the length of the screw, gives the height of the column.

Other forms of the instrument are used, most of which are arranged with reference to convenient transportability. Various contrivances are added by means of which the column is made to move an index, and thus record the pressure on a graduated scale. All these forms are only modifications of Torricelli's original instrument.

The pressure indicated by the barometer is usually stated in terms of the height of the column. Mercury being practically incompressible, this height is manifestly proportional to the pressure at any point in the surface of the mercury in the cistern. The pressure on any given area in that surface can be calculated if we know the value of g at the place and the specific gravity of mercury, as well as the height of the column. The standard barometric pressure, represented by 760 millimetres of mercury, is a pressure of 1.033 kilograms on every square centimetre. It is called a pressure of one *atmosphere*; and pressures are often measured by atmospheres.

In the preparation of an accurate barometer it is necessary that all air be removed from the mercury; otherwise it will collect in the upper part of the tube, by its pressure lower the top of the column, and make the barometer read too low. The air is removed by partially filling the tube with mercury, which is then boiled in the tube, gradually adding small quantities of mercury, and boiling

after each addition, until the tube is filled. The boiling must not be carried too far; for there is danger, in this process, of expelling the air so completely that the mercury will adhere to the sides of the tube, and will not move freely. For rough work the tube may be filled with cold mercury, and the air removed by gently tapping the tube, so inclining it that the small bubbles of air which form can coalesce, and finally be set free at the surface of the mercury.

115. Archimedes' Principle.—If a solid be immersed in a fluid, it loses in weight an amount equal to the weight of the fluid displaced. This law is known, from its discoverer, as *Archimedes' principle.*

The truth of this law will appear if we consider the space in the fluid which is afterwards occupied by the solid. The fluid in this space will be in equilibrium, and the upward pressure on it must exceed the downward pressure by an amount equal to its weight. The resultant of the pressure acts through the centre of gravity of the assumed portion of fluid, otherwise equilibrium would not exist. If, now, the solid occupy the space, the difference between the upward and the downward pressures on it must still be the same as before,—namely, the weight of the fluid displaced by the solid; that is, the solid loses in apparent weight an amount equal to the weight of the displaced fluid.

116. Floating Bodies.—When the solid floats on the fluid, the weight of the solid is balanced by the upward pressure. In order that the solid shall be in equilibrium, these forces must act in the same line. The resultant of the pressure, which lies in the vertical line passing through the centre of gravity of the displaced fluid, must pass through the centre of gravity of the solid. Draw the line in the solid joining these two centres, and call it the *axis* of the solid. The equilibrium is stable when, for any infinitesimal inclination of the axis from the vertical, the vertical line of upward pressure cuts the axis in a point above the centre of gravity of the solid. This point is called the *metacentre.*

117. Specific Gravity.—Archimedes' principle is used to determine the specific gravity of bodies. The *specific gravity* of a body

is defined as the ratio of its weight to the weight of an equal volume of pure water at a standard temperature.

The specific gravity of a solid that is not acted on by water may be determined by means of the *hydrostatic balance*. The body under examination, if it will sink in water, is suspended from one scale-pan of a balance by a fine thread, and is weighed. It is then immersed in water, and is weighed again. The difference between the weights in air and in water is the weight of the displaced water, and the ratio of the weight of the body to the weight of the displaced water is the specific gravity of the body.

If the body will not sink in water, a sinker of unknown weight and specific gravity is suspended from the balance, and counterpoised in water. Then the body, the specific gravity of which is sought, is attached to the sinker, and it is found that the equilibrium is destroyed. To restore it, weights must be added to the same side. These, being added to the weight of the body, represent the weight of the water displaced.

The specific gravity of a liquid is obtained by first balancing in air a mass of some solid, such as platinum or glass, that is not acted on chemically by the liquid, and then immersing the mass successively in the liquid to be tested and in water. The ratio of the weights which must be used to restore equilibrium in each case is the specific gravity of the liquid.

The specific gravity of a liquid may also be found by means of the *specific gravity bottle*. This is a bottle fitted with a ground-glass stopper. The weight of the water which completely fills it is determined once for all. When the specific gravity of any liquid is desired, the bottle is filled with the liquid, and the weight of the liquid determined. The ratio of this weight to the weight of an equal volume of water is the specific gravity of the liquid.

The same bottle may be used to determine the specific gravity of any solid which cannot be obtained in continuous masses, but is friable or granular. A weighed amount of the solid is introduced into the bottle, which is then filled with water, and the weight of the joint contents of the bottle determined. The difference

between the last weight and the sum of the weights of the solid and of the water filling the bottle is the weight of the water displaced by the solid. The ratio of the weight of the solid to the weight thus obtained is the specific gravity of the solid.

The specific gravity of a liquid may also be obtained by means of *hydrometers*. These are of two kinds—the hydrometers of *constant weight* and those of *constant volume*. The first consists usually of a glass bulb surmounted by a cylindrical stem. The bulb is weighted, so as to sink in pure water to some definite point on the stem. This point is taken as the zero; and, by successive trials with different liquids of known specific gravity, points are found on the stem to which the hydrometer sinks in these liquids. With these as a basis, the divisions of the scale are determined and cut on the stem.

The hydrometer of constant volume consists of a bulb weighted so as to stand upright in the liquid, bearing on the top of a narrow stem a small pan, in which weights may be placed. The weight of the hydrometer being known, it is immersed in water; and, by the addition of weights in the pan, a fixed point on the stem is brought to coincide with the surface of the water. The instrument is then transferred to the liquid to be tested, and the weights in the pan changed until the fixed point again comes to the surface of the liquid. The sum of the weight of the hydrometer and the weights added in each case gives the weight of equal volumes of water and of the liquid, from which the specific gravity sought is easily obtained.

The specific gravity of gases is often referred to air or to hydrogen instead of water. It is best determined by filling a large glass flask, of known weight, with the gas, the specific gravity of which is to be obtained, and weighing it, noting the temperature and the pressure of the gas in the flask. The weight of the gas at the standard temperature and pressure is then calculated, and the ratio of this weight to the weight of the same volume of the standard gas is the specific gravity desired. The weight of the flask used in this experiment must be very exactly determined. The presence

of the air vitiates all weighings performed in it, by diminishing the true weight of the body to be weighed and of the weights employed, by an amount proportional to their volumes. The consequent error is avoided either by performing the weighings in a vacuum produced by the air-pump, or by correcting the apparent weight in air to the true weight. Knowing the specific gravity of the weights and of the body to be weighed, and the specific gravity of air, this can easily be done.

118. Motions of Fluids.—If the parts of the fluid be moving relatively to each other or to its bounding-surface, the circumstances of the motion can be determined only by making limitations which are not actually found in Nature. There thus arise certain definitions to which we assume that the fluid under consideration conforms.

The motion of a fluid is said to be *uniform* when each element of it has the same velocity at all points of its path. The motion is *steady* when, at any one point, the velocity and direction of motion of the elements successively arriving at that point remain the same for each element. If either the velocity or direction of motion change for successive elements, the motion is said to be *varying.* The motion is further said to be *rotational* or *irrotational* according as the elements of the fluid have or have not an angular velocity about their axes.

In all discussions of the motions of fluids a condition is supposed to hold, called the *condition of continuity.* It is assumed that, in any volume selected in the fluid, the change of density in that volume depends solely on the difference between the amounts of fluid flowing into and out of that volume. In an incompressible fluid, or liquid, if the influx be reckoned plus and the efflux minus, we have, letting Q represent the amount of the liquid passing through the boundary in any one direction, $\Sigma Q = 0$. The results obtained in the discussion of fluid motions must all be interpreted consistently with this condition. If the motion be such that the fluid breaks up into discontinuous parts, any results obtained by hydrodynamical considerations no longer hold true.

If we consider any stream of incompressible fluid, of which the cross-sections at two points where the velocities of the elements are v_1 and v_2 have respectively the areas A_1 and A_2, we can deduce at once from the condition of continuity

$$A_1v_1 = A_2v_2. \tag{52}$$

119. Velocity of Efflux.—We shall now apply this principle to discover the *velocity of efflux* of a liquid from an orifice in the walls of a vessel.

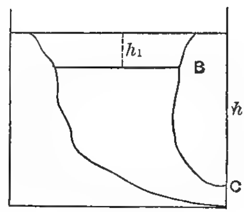

FIG. 39.

Consider any small portion of the liquid, bounded by stream lines, which we may call a *filament*. Represent the velocity of the filament at B (Fig. 39) by v_1, and at C by v, and the areas of the cross-sections of the filament at the same points by A_1 and A. We have then, as above, $A_1v_1 = Av$. We assume that the flow has been established for a time sufficiently long for the motion to become steady. The energy of the mass contained in the filament between B and C is, therefore, constant. Let V_1 represent the potential or the potential energy of unit mass at B due to gravity, V the potential at C, and d the density of the liquid. The mass that enters at B in a unit of time or the rate at which mass enters at B is dA_1v_1. The rate at which mass goes out at C is the equal quantity dAv. The energy entering at B is $dA_1v_1(\frac{1}{2}v_1^2 + V_1)$, the energy passing out at C is $dAv(\frac{1}{2}v^2 + V)$.

If the pressures at B and C on unit areas be expressed by p_1 and p, the rate at which work is done at B on the entering mass by the pressure p_1 is $p_1A_1v_1$, and at C on the outgoing mass is pAv. This may be seen by considering the cross-section of the filament at C. The pressure p acting on each unit of area of that cross-section is equivalent to a force pA, and v is the rate at which the cross-section moves forward, so that pAv is the rate at which the pressure does work. The energy within the filament remaining constant, the incoming must equal the outgoing energy; therefore

$pAv + dAv(\frac{1}{2}v^2 + V) = p_1A_1v_1 + dA_1v_1(\frac{1}{2}v_1^2 + V_1)$, whence, since $A_1v_1 = Av$, we have $\frac{p}{d} + \frac{1}{2}v^2 + V = \frac{p_1}{d} + \frac{1}{2}v_1^2 + V_1$.

By using again the relation $A_1v_1 = Av$, this equation becomes

$$\frac{1}{2}v^2\left(1 - \frac{A^2}{A_1^2}\right) = (V_1 - V) + \frac{p_1 - p}{d}. \tag{53}$$

To apply equation (53) to the case of a liquid flowing freely into air from an orifice at C, we observe that the difference of potential $(V_1 - V)$ equals the work done in carrying a gram from C to B or equals $g(h - h_1)$, where h represents the height of the surface above C, and h_1 that of the surface above B. Further we have $p_1 = p_a + dgh_1$, where p_a is the atmospheric pressure. At the orifice p equals p_a. We have then $\frac{1}{2}v^2\left(1 - \frac{A^2}{A_1^2}\right) = g(h - h_1) + gh_1 = gh$, whence $v^2 = \dfrac{2gh}{1 - \dfrac{A^2}{A_1^2}}$. If, now, A becomes indefinitely small as compared with A_1, in the limit the velocity at C becomes

$$v = \sqrt{2gh}\,; \tag{54}$$

that is, the velocity of efflux of a small stream issuing from an orifice in the wall of a vessel is independent of the density of the liquid, and is equal to the velocity which a body would acquire in falling freely through a distance equal to that between the surface of the liquid and the orifice.

This theorem was first given by Torricelli from considerations based on experiment, and is known as *Torricelli's theorem.* Its demonstration is due to Daniel Bernoulli.

We may apply the general equation to the case of the efflux of a liquid through a siphon. A *siphon* is a bent tube which is used to convey a liquid by its own weight over a barrier. One end of the siphon is immersed in the liquid, and the discharging end, which must be below the level of the liquid, opens on the other side of

the barrier. To set the siphon in operation it must be first filled with the liquid, after which a steady flow is maintained.

In this case, as before, we may set $\frac{A^2}{A_1^2} = 0$, $v_1 = 0$, p and p_1 both $= p_a$, and $(V_1 - V) = gl$, where l is the distance between the surface level and the discharging orifice. The velocity becomes $v = \sqrt{2gl}$. The siphon, therefore, discharges more rapidly the greater the distance between the surface level and the orifice. It is manifest that the height of the bend in the tube cannot be greater than that at which atmospheric pressure would support the liquid.

The flow of a liquid into the vacuum formed in the tube of an ordinary pump may also be discussed by the same equation. The *pump* consists essentially of a tube, fitted near the bottom with a partition, in which is a valve opening upwards. In the tube slides a tightly fitting piston, in which is a valve, also opening upwards. The piston is first driven down to the partition in the tube, and the enclosed air escapes through the valve in the piston. When the piston is raised, the liquid in which the lower end of the tube is immersed passes through the valve in the partition, rises in the tube and fills the space left behind the piston. When the piston is again lowered, the space above it is filled with the liquid, which is lifted out of the tube at the next up-stroke.

To determine the velocity of the liquid following the piston, we notice that in this case $p_1 = p_a$ and $p = 0$ if the piston move upward very rapidly, $(V_1 - V) = -gh$, where h is the height of the top of the liquid column above the free surface in the reservoir, and $\frac{A^2}{A_1^2}$ again $= 0$. We then have $\frac{1}{2}v^2 = \frac{p_a}{d} - gh$.

The velocity when $h = 0$ is $v = \sqrt{\frac{2p_a}{d}}$. When h is such that $dgh = p_a$, $v = 0$, which expresses the condition of equilibrium.

The equation $v = \sqrt{\frac{2p_a}{d}}$ expresses, more generally, the velocity

of efflux, through a small orifice, of any fluid of density d, from a region in which it is under a constant pressure p_a, into a vacuum.

Torricelli's theorem is shown to be approximately true by allowing liquids to run from an orifice in the side of a vessel, and measuring the path of the stream. If the theorem be true, this ought to be a parabola, of which the intersection of the plane of the stream and of the surface of the liquid is the directrix; for each portion of the liquid, after it has passed the orifice, will behave as a solid body, and move in a parabolic path. The equation of this path is found, as in § 52, to be $x^2 = -\frac{2v^2}{g} y$. Now by Torricelli's theorem we may substitute for v^2 its value $2gh$, whence $x^2 = -4hy$. In this equation, since the initial movement of the stream is supposed to be horizontal, the perpendicular line through the orifice being the axis of the parabola, and the orifice being the origin, h is the distance from the orifice to the directrix. Experiments of this kind have been frequently tried, and the results found to approximate more nearly to the theoretical as various causes of error were removed.

When, however, we attempt to calculate the amount of liquid discharged in a given time, there is found to be a wider discrepancy between the results of calculation and the observed facts. Newton first noticed that the diameter of the jet at a short distance from the orifice is less than that of the orifice. He showed this to be a consequence of the freedom of motion among the particles in the vessel. The particles flow from all directions towards the orifice, those moving from the sides necessarily issuing in streams inclined towards the axis of the jet. Newton showed that by taking the diameter of the narrow part of the jet, which is called the *vena contracta*, as the diameter of the orifice, the calculated amount of liquid escaping agreed far more closely with theory.

When the orifice is fitted with a short cylindrical tube, the interference of the different particles of the liquid is in some degree lessened, and the quantity discharged increases nearly to that required by theory.

120. Diminution of Pressure.—The Sprengel air-pump, an important piece of apparatus to be described hereafter, depends for its operation on the diminution of pressure at points along the line of a flowing column of liquid. Let us consider a large reservoir filled with liquid, which runs from it by a vertical tube entering the bottom of the reservoir. From equation (53) the value of p, the pressure at any point in the tube, is $p = p_1 + \ (V_1 - V)d - \frac{1}{2}dv^2\left(1 - \frac{A^2}{A_1}\right)$. The ratio $\frac{A^2}{A_1^2}$ may be set equal to zero. If h (Fig. 40) represent the height of the upper surface above the point in the tube at which we desire to find the pressure, then $(V_1 - V) = gh$. We then have $p = p_a + dgh - \frac{1}{2}dv^2$. If the tube be always filled with the liquid, $Av = A_0v_0$, where A and A_0 represent the areas of the cross-sections of the tube at the point we are considering and at the bottom of the tube, and v and v_0 represent the corresponding velocities. Further, $v_0^2 = 2gh_0$ if h_0 represent the distance from the upper surface to the bottom of the tube. We obtain, by substitution,

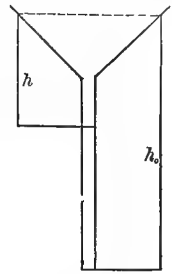

FIG. 40.

$$p = p_a + dg\left(h - \frac{A_0^2}{A^2}h_0\right). \tag{55}$$

If h equal $\frac{A_0^2}{A^2}h_0$, we have $p = p_a$; and if an opening be made in the wall of the tube, the moving liquid and the air will be in equilibrium. If h be less than $\frac{A_0^2}{A^2}h_0$, the pressure p will be less than p_a, and air will flow into the tube. Since this inequality exists when $A_0 = A$, it follows that, if a liquid flow from a reservoir down a cylindrical tube, the pressure at any point in the wall of the tube is less than the atmospheric pressure by an amount equal to the pressure of a column of the liquid, the height of which is equal to the distance between the point considered and the bottom of the tube.

121. Waves.—When a disturbance is set up at a point in the free surface of a liquid, it moves over the surface of the liquid as a *wave* or series of waves. Each wave consists of a crest or elevated portion and a hollow or depressed portion of approximately equal length, and the distance from a particle at the summit of one crest to a particle at the summit of the next succeeding crest, or the distance between particles in successive waves which are in the same condition of motion, is called a *wave length.* A line which is drawn along the crest of any one wave or through the particles in that wave which are in the same condition of motion, and which at every point is at right angles to the direction in which the wave is propagated, may be called the *wave front.*

The formation of waves is explained by inequalities of hydrostatic pressure arising in the liquid if by any cause one part of it be elevated above the rest. H. and W. Weber examined the peculiarities of waves in water and the motions of the water particles in them by the aid of a long trough with glass sides; by immersing one end of a glass tube below the surface, raising a column of water in it a few centimetres high by suction, and allowing it to fall, they excited a series of waves which proceeded down the trough and could be examined through the sides. The motions of the particles in the wave were studied by scattering through the water small fragments of amber, which were so nearly of the same specific gravity as the water that they remained suspended without motion except during the passage of the wave, and took part in the motion excited by the wave as if they had been particles of water. It was found that the wave motion was a form of motion transferred from one portion to another of the water, and did not involve a displacement of the particles concerned in it,—at least when the successive waves had the same wave length. In that case—which is the typical one—the particles in the surface of the water described closed curves, which were elliptical or circular in form, the diameter of the circle being equal to the vertical distance between the crest and the hollow or the *height* of the wave. In the upper part of the circle the particle moved in the direction in which the wave was

moving, in the lower part of the circle in the opposite direction. The velocity of the wave was found to be dependent on its height and on the period of oscillation in the wave, and to be independent of the density of the liquid. The disturbance of the liquid by the wave is not merely on the surface, but extends to a considerable depth; as the depth increases the elliptic paths of the particles approach more and more closely to short horizontal lines.

The theory of these waves is extremely complicated, and has not yet been satisfactorily worked out; but we can indicate in a general way their causes and the mode of their propagation. Imagine a small hillock of water elevated at some point in the surface, and consider a particle at the base of this hillock; the hydrostatic pressure arising from the elevated column near it will tend to move it upward and outward from the centre of the hillock. It will accordingly begin to move in the upper half of its circular path and in the direction in which the wave is propagated; the precise form of its path being determined by the changes of pressure which it experiences and by its inertia. Since the pressure which sets it in motion will be different for different heights of the hillock which gives rise to it, the velocity of the particles, and therefore also the velocity of the wave, will depend on the height of the wave, being greater as this is greater; the velocity of the wave is also greater as the wave length is greater. Since the pressure behind the particle and the inertia are both proportional to the density of the liquid, it is evident that the acceleration of the particle will be the same under similar circumstances, whatever be its density, so that the velocity of the wave should not depend on the density of the liquid.

The form of a wave is greatly modified by the character of the channel in which it moves, on account of the motion of the particles extending to a considerable depth, and on account of their viscosity. On the free surface of a large and very deep body of water the successive waves have the same form; the slope of the crest is a little steeper than the slope of the hollow, and its length is less than that of the hollow. As the depth decreases, the slope

of the front of the crest becomes still steeper because of the restraint which then is imposed upon the movement of the particles in the lower half of their paths, and at last the forward motion in the crest so much predominates that the wave curls over and breaks.

122. Vortices.—A series of most interesting results has been obtained by Helmholtz, Thomson, and others, from the discussion of the rotational motions of fluids. Though the proofs are of such a nature that they cannot be presented here, the results are so important that they will be briefly stated.

A *vortex line* is defined as the line which coincides at every point with the instantaneous axis of rotation of the fluid element at that point. A *vortex filament* is any portion of the fluid bounded by vortex lines.

A *vortex* is a vortex filament which has "contiguous to it over its whole boundary irrotationally moving fluid."

The theorems relating to this form of motion, as first proved by Helmholtz, in 1868, show that,—

(1) A vortex in a perfect fluid always contains the same fluid elements, no matter what its motion through the surrounding fluid may be.

(2) The *strength* of a vortex, which is the product of its angular velocity by its cross-section, is constant; therefore the vortex in an infinite fluid must always be a closed curve, which, however, may be knotted and twisted in any way whatever.

(3) In a finite fluid the vortex may be open, its two ends terminating in the surface of the fluid.

(4) The irrotationally moving fluid around a vortex has a motion due to its presence, and transmits the influence of the motion of one vortex to another.

(5) If the vortices considered be infinitely long and rectilinear, any one of them, if alone in the fluid, will remain fixed in position.

(6) If two such vortices be present parallel to one another, they revolve about their common centre of mass.

(7) If the vortices be circular, any one of them, if alone, moves with a constant velocity along its axis, at right angles to the plane

of the circle, in the direction of the motion of the fluid rotating on the inner surface of the ring.

(8) The fluid encircled by the ring moves along its axis in the direction of the motion of the ring, and with a greater velocity.

(9) If two circular vortices move along the same axis, one following the other, the one in the rear moves faster, and diminishes in diameter; the one in advance moves slower, and increases in diameter. If the strength and size of the two be nearly equal, the one in the rear overtakes the other, and passes through it. The two now having changed places, the action is repeated indefinitely.

(10) If two circular vortices of equal strength move along the same axis toward one another, the velocities of both gradually decrease and their diameters increase. The same result follows if one such vortex move toward a solid barrier.

The preceding statements apply only to vortices set up in a perfect fluid. They may, however, be illustrated by experiment. To produce circular vortices in the air, we use a box which has one of its ends flexible. A circular opening is cut in the opposite end. The box is filled with smoke or with finely divided sal-ammoniac, resulting from the combination of the vapors of ammonia and hydrochloric acid. On striking the flexible end of the box smoke-rings are at once sent out.

The smoke-ring is easily seen to be made up of particles revolving about a central core in the form of a ring. With such rings many of the preceding statements may be verified.

An illustration of the open vortex is seen when an oar-blade is drawn through the water. By making such open vortices, using a circular disk, many of the observations with the smoke-rings may be repeated in another form.

123. Air-pumps.—The fact that gases, unlike liquids, are easily compressed, and obey Boyle's law under ordinary conditions of temperature and pressure, underlies the construction and operation of several pieces of apparatus employed in physical investigations. The most important of these is the *air-pump*.

The working portion of the air-pump is constructed essentially like the common lifting-pump already described. The valves must be light and accurately fitted. The vessel from which the air is to be exhausted is joined to the pump by a tube, the orifice of which is closed by the valve in the bottom of the cylinder.

A special form of vessel much used in connection with the air-pump is called the *receiver*. It is usually a glass cylinder, open at one end and closed by a hemispherical portion at the other. The edge of the cylinder at the open end is ground perfectly true, so that all points in it are in the same plane. This ground edge fits upon a plane surface of roughened brass, or ground glass, called the *plate*, through which enters the tube which joins the receiver to the cylinder of the pump. The joint between the receiver and the plate is made tight by a little oil or vaseline.

The action of the pump is as follows: As the piston is raised, the pressure on the upper surface of the valve in the cylinder is diminished, and the air in the vessel expands in accordance with Boyle's law, lifts the valve, and distributes itself in the cylinder, so that the pressure at all points in the vessel and the cylinder is the same. The piston is now forced down, the lower valve is closed by the increased pressure on its upper surface, the valve in the piston is opened, and the air in the cylinder escapes. At each successive stroke of the pump this process is repeated, until the pressure of the remnant of air left in the vessel is no longer sufficient to lift the valves.

The density of the air left in the vessel after a given number of strokes is determined, provided there be no leakage, by the relations of the volumes of the vessel and the cylinder.

Let V represent the volume of the vessel, and C that of the cylinder when the piston is raised to the full extent of the stroke. Let d and d_1 respectively represent the density of the air in the vessel before and after one stroke has been made. After one down and one up stroke have been made, the air which filled the volume V now fills $V + C$. It follows that $\frac{d_1}{d} = \frac{V}{V + C}$. As this ratio is

constant no matter what density may be considered, it follows that, if d_n represent the density after n strokes,

$$\frac{d_n}{d} = \left(\frac{V}{V + C}\right)^n. \tag{56}$$

As this fraction cannot vanish until n becomes infinite, it is plain that a perfect vacuum can never, even theoretically, be obtained by means of the air-pump. If, however, the cylinder be large, the fraction decreases rapidly, and a few strokes are sufficient to bring the density to such a point that either the pressure is insufficient to lift the valves, or the leakage through the various joints of the pump counterbalances the effect of longer pumping.

In the best air pumps the valves are made to open automati-

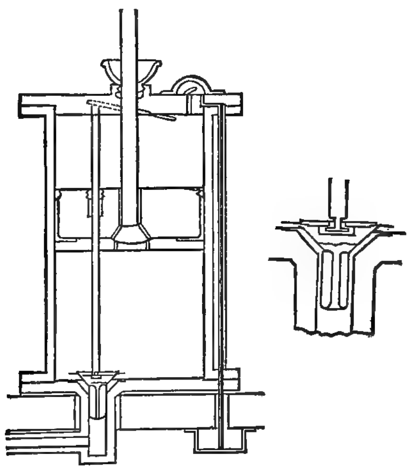

FIG. 41.

cally. In Fig. 41 is represented one of the methods by which this is accomplished. They can then be made heavier and with a larger surface of contact, so that the leakage is diminished, and the limit of the useful action of the pump is much extended. With the best pumps of this sort a pressure of half a millimetre of mercury is reached.

The *Sprengel air-pump* depends for its action upon the princi-

ple, discussed in § 120, that a stream of liquid running down a cylinder diminishes the pressure upon its walls. In the Sprengel pump the liquid used is mercury. It runs from a large vessel down a glass tube, into the wall of which, at a distance from the bottom of the tube of more than 760 millimetres, enters the tube which connects with the receiver. The lower end of the vertical tube dips into mercury, which prevents air from passing up along the walls of the tube. When the stream of mercury first begins to flow, the air enters the column from the receiver, in consequence of the diminished pressure, passes down with the mercury in large bubbles, and emerges at the bottom of the tube. As the exhaustion proceeds, the bubbles become smaller and less frequent, and the mercury falls in the tube with a sharp, metallic sound. It is evident that, as in the case of the ordinary air-pump, a perfect vacuum cannot be secured. There is no leakage, however, in this form of the air-pump, and a very high degree of exhaustion can be reached.

The Morren or Alvergniat *mercury-pump* is in principle merely a common air-pump, in which combinations of stop-cocks are used instead of valves, and a column of mercury in place of the piston. Its particular excellence is that there is scarcely any leakage.

The *compressing-pump* is used, as its name implies, to increase the density of air or any other gas within the receiver. The receiver in this case is generally a strong metallic vessel. The working parts of the pump are precisely those of the air-pump, with the exception that the valves open downwards. As the piston is raised, air enters the cylinder, and is forced into the receiver at the down-stroke.

124. Manometers.—The *manometer* is an instrument used for measuring pressures. One variety depends for its operation upon the regularity of change of volume of a gas with change of pressure. This, in its typical form, consists of a heavy glass tube of uniform bore, closed at one end, with the open end immersed in a basin of mercury. The pressure to be measured is applied to the surface of the mercury in the basin. As this pressure increases, the air contained in the tube is compressed, and a column of mer-

cury is forced up the tube. The top of this column serves as an index. We know from Boyle's law that when the volume of the air has diminished one half, the pressure is doubled. The downward pressure of the mercury column makes up a part of this pressure; and the pressure acting on the surface of the mercury in the basin is greater than that indicated by the compression of the air in the tube, by the pressure due to the mercury column. For many purposes the manometer tube may be made very short, and the pressure of the mercury column that rises in it may be neglected.

125. Aneroids.—The *aneroid* is an instrument used to determine ordinary atmospheric pressures. On account both of its delicacy and its easy transportability, it is often used instead of the barometer. It consists of a metallic box, the cover of which is made of thin sheet metal corrugated in circular grooves. The air is partially exhausted from the box, and it is then sealed. Any change in the pressure of the atmosphere causes the corrugated top to move. This motion is very slight, but is made perceptible, either by a combination of levers, which amplifies it, or by an arm rigidly fixed on the top, the motion of which is observed by a microscope. The indications of the aneroid are compared with those of a standard mercurial barometer, and an empirical scale is thus made, by means of which the aneroid may be used to determine pressures directly.

126. Limitations to the Accuracy of Boyle's Law.—In all the previous discussions we have dealt with gases as if they obeyed Boyle's law with absolute exactness. This, however, is not the case. In the first place, some gases at ordinary temperatures can be liquefied by pressure. As these gases approach more nearly the point of liquefaction, the product pv of the volume and pressure becomes less than it ought to be in accordance with Boyle's law.

Secondly, those gases which cannot be liquefied at ordinary temperatures by any pressure, however great, show a different departure from the law. For every gas, except hydrogen, there is a minimum value of the product pv. At ordinary temperatures and small pres-

sures the gas follows Boyle's law quite closely, becoming, however, more compressible as the pressure increases, until the minimum value of pv is reached. It then becomes gradually less compressible, and at high pressures its volume is much greater than that determined by Boyle's law. If the temperature be raised, the agreement with the law is closer, and the pressure at which the minimum value of pv occurs is greater. Hydrogen seems to differ from the other gases, only in that the pressures at which the observations upon it were made were probably greater than the one at which its minimum value of pv occurs. The volume of the compressed hydrogen is uniformly greater than that required by Boyle's law.

Important modifications are introduced into the behavior of gases under pressure by subjecting them to intense cold. It is then found that all gases, without exception, can be liquefied, and most of them solidified.

The subject is intimately connected with the subject of critical temperature, and will be again discussed under Heat.

SOUND.

CHAPTER I.

ORIGIN AND TRANSMISSION OF SOUND.

127. Definitions.—*Acoustics* has for its object the study of those phenomena which may be perceived by the ear. The sensations produced through the ear, and the causes that give rise to them, are called *sounds*.

128. Origin of Sound.—Sound is produced by vibratory movements in elastic bodies. The vibratory motion of bodies when producing sound is often evident to the eye. In some cases the sound seems to result from a continuous movement, but even in these cases the vibratory motion can be shown by means of an apparatus known as a *manometric capsule*, devised by König. It consists of a block A (Fig. 42) in which is a cavity covered by a membrane b. By means of a tube c illuminating gas is led into the cavity, and, passing out through the tube d, burns in a jet at e. It is evident that, if the membrane b be made to move suddenly inward or outward, it will compress or

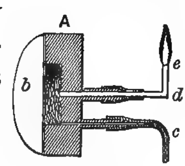

Fig. 42.

rarefy the gas in the capsule, and so cause the flow at the orifice and the height of the flame to increase or diminish. Any sound of sufficient intensity in the vicinity of the capsule causes an alternate lengthening and shortening of the flame, which, however, occurs too frequently to be directly observed. By moving the eyes while keeping the flame in view, or by observing the image of the

flame in a mirror which is turned from side to side, while the flame is quiescent, it appears drawn out into a broad band of light, but when it is agitated by a sound near it, it appears serrate on its upper edge or even as a series of separate flames. This lengthening and shortening of the flame is evidence of a to-and-fro movement of the membrane, and hence of the sounding body that gave rise to the movement. If a hole be made in the side of an organ-pipe and the capsule made to cover it, the vibrations of the air-column within the pipe may be shown. By suitable devices the vibratory motion of all sounding bodies may be demonstrated.

129. Propagation of Sound.—The vibratory motion of a sounding body is ordinarily transmitted to the ear through the air. This is proved by placing a sounding body under the receiver of an air-pump and exhausting the air. The sound becomes fainter and fainter as the exhaustion proceeds, and finally becomes inaudible if the vacuum is good. Sound may, however, be transmitted by any elastic body.

In order to study the character of the motion by which sound is propagated, let us suppose AB (Fig. 43) to represent a cylinder

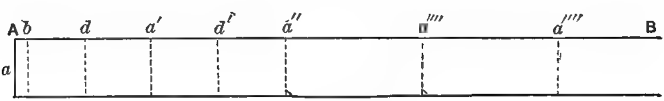

FIG. 43.

of some elastic substance, and suppose the layer of particles a to suffer a small displacement to the right. The effect of this displacement is not immediately to move forward the succeeding layers, but a approaches b, producing a condensation, and developing a force that soon moves b forward; this in turn moves forward the next layer, and so the motion is transmitted from layer to layer through the cylinder with a velocity that depends upon the elasticity (§ 103) of the substance, and upon its density. This velocity is expressed by the formula $V = \sqrt{\frac{E}{D}}$, in which E represents the elasticity of the substance, and D its density (§ 134). Now, if we

suppose the layer a, from any cause whatever, to execute regular vibrations, this movement will be transmitted to the succeeding layers with the velocity given by the formula, and, in time, each layer of particles in the cylinder will be executing vibrations similar to those of a. If the vibrations of a be performed in the time t, the motion will be transmitted during one complete vibration of a to a distance $s = vt$, where v is the velocity of propagation, say to a', during two complete vibrations of a to a distance $2s = 2vt$, or to a'', during three complete vibrations to a''', and so on. It is evident that the layer a' begins its first vibration at the instant that a begins its second vibration, a'' begins its first vibration at the instant that a' begins its second and a its third vibration. The layer midway between a and a' evidently begins its vibration just as a completes the first half of its vibration, and therefore moves forward while a moves backward. This condition of things existing in the cylinder constitutes a *wave motion*. While a moves forward, the portions near it are compressed. While it moves backward, they are dilated. Whatever the condition at a, the same condition will exist at the same instant at a', a'', etc. The distance $aa' = a'a''$ is called a *wave length;* it is the distance from any one particle to the next one of which the vibrations are in the same phase (§ 21). If the condition at a and a' be one of *condensation*, it is evident that at d, midway between a and a', there must be a *rarefaction*. In the wave length aa' exist all intermediate conditions of condensation and rarefaction. These conditions must follow each other along the cylinder with the velocity of the transmitted motion, and they constitute a *progressive wave* moving with this velocity. If the vibratory motion with which a is endowed be communicated by a sounding body, the wave is a *sound-wave*. If, instead of a cylinder of the substance, we have an indefinite medium in the midst of which the sounding body is placed, the motion is transmitted in all directions as spherical waves about the sounding body as a centre.

130. Mode of Propagation of Wave Motion.—The mode of transmission of wave motions was first shown by Huygens, and

the principle involved is known as *Huygens' principle.* Let a (Fig. 44) be a centre at which sound originates. At the end of a certain time it will have reached the surface mn. From the preceding discussion it is evident that each particle of the surface mn has a vibratory motion similar to that at a. Any one of those particles would, if vibrating alone, be, like a, the centre of a system of spherical waves, and each of them must, therefore, be considered as a wave centre from which spherical waves proceed. Suppose such a wave to proceed from each one of them for the short distance cd. Since the number of the elementary spherical waves is very great, it is plain that they will coalesce to form the surface $m'n'$ which determines a new position of the wave surface. In some cases the existence of these elementary waves need not be considered, but there are many phenomena of wave motion which can only be studied by recognizing the fact that propagation always takes place as above described.

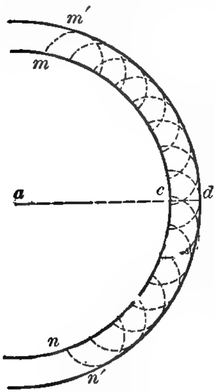

FIG. 44.

131. Graphic Representation of Wave Motion.—In order to study the movements of a body in which a wave motion exists, especially when two or more systems of waves exist in the same body, it is convenient to represent the movement by a sinusoidal curve.

Suppose the layer a (Fig. 45) to move with a simple harmonic motion of which the amplitude is a and the period T, and let time be reckoned from the instant that the particles pass the position of equilibrium in a positive direction. A sinusoidal curve may be constructed to represent either the displacements of the various layers from their positions of equilibrium, or the velocities with which they are severally moving at a given time.

To construct the first curve let the several points along OX (Fig. 45) represent points of the body through which the wave is moving. Let $Oy = a$ be the amplitude of vibration of each particle. The displacement of the particle at O at any instant t after

passing its position of equilibrium is $y = a \cos\left(\frac{2\pi t}{T} - \frac{\pi}{2}\right)$, since when t is reckoned from the position of equilibrium $\epsilon = \frac{\pi}{2}$. Hence $y = a \sin \frac{2\pi t}{T}$. If v represent the velocity of propagation of the wave, the particle at the distance x from the origin will have a displacement equal to that of the particle at O at the instant t, at an

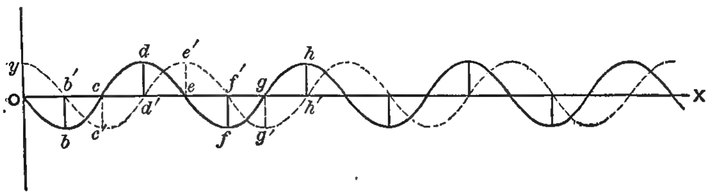

FIG. 45.

instant later than t by the time taken for the wave to travel over the distance x, or $\frac{x}{v}$ seconds. Hence its displacement at the instant t will be the same as that which existed at O, $\frac{x}{v}$ seconds earlier. But the displacement at O, $\frac{x}{v}$ seconds earlier, is

$$y = a \sin \frac{2\pi\left(t - \frac{x}{v}\right)}{T} = a \sin 2\pi\left(\frac{t}{T} - \frac{x}{vT}\right). \tag{57}$$

The quantity vT equals the distance through which the movement is transmitted during the time of one complete vibration of the particle at O. Putting this equal to λ, we have finally

$$y = a \sin 2\pi\left(\frac{t}{T} - \frac{x}{\lambda}\right). \tag{58}$$

Suppose $t = 0$, and give to x various values. The corresponding values of y will represent the displacement at that instant of the particle the distance of which from the origin is x. For $x = 0$,

$y = 0$. For $x = \frac{1}{4}\lambda$, $y = -a$. For $x = \frac{1}{2}\lambda$, $y = 0$. For $x = \frac{3}{4}\lambda$, $y = a$. For $x = \lambda$, $y = 0$, etc. Laying off these values of x on OX and erecting perpendiculars equal to the corresponding values of y, we have the curve $Obcde$ . . . .

The above expression for y may be put in the form

$$y = a \sin 2\pi \left(\frac{\frac{t\lambda}{T} - x}{\lambda} \right).$$

Hence, if any finite value be assigned to t, we shall obtain for y the same values as were obtained above for $t = 0$, if we increase each of the values of x by $\frac{t\lambda}{T}$. For instance, if t equal $\frac{1}{4}T$, we have $y = 0$ for $x = \frac{1}{4}\lambda$, $y = -a$ for $x = \frac{1}{2}\lambda$, etc., and the curve becomes the dotted line $b'c'd'$ . . . . The effect of increasing t is to displace the curve along OX in the direction of propagation of the wave.

The formula for constructing the curve of velocities is derived in the same way as that for displacements. It is

$$y = \frac{2\pi a}{T} \cos 2\pi \left(\frac{t}{T} - \frac{x}{\lambda} \right).$$

Fig. 46 shows the relation of the two curves. The upper is the curve of displacement, and the lower of velocity.

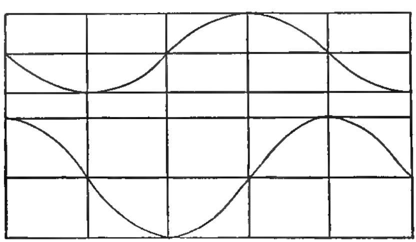

FIG. 46.

132. Composition of Wave Motions.—The composition of wave motions may be studied by the help of the curves explained above. If two systems of waves coexist in the same body, the displacement of any particle at any instant will be the algebraic sum of the dis-

placements due to the systems taken separately. If the curve of displacements be drawn for each system, the algebraic sum of the ordinates will give the ordinates of the curve representing the actual displacements. In Fig. 47 the dotted line and the light full line represent respectively the displacements due to two wave systems of the same period and amplitude. The heavy line represents the actual displacement. In I the two systems are in the same phase; in II the phases differ by $\frac{1}{4}$, and in III by $\frac{1}{2}$, of a period. If both wave systems move in the same direction, it is evident that the conditions of the body will be continuously shown by supposing the heavy line to move in the same direction with the same velocity. The condition represented in III is of special interest. It shows that two wave systems may completely annul each other. Fig. 48 represents the resultant wave when the pe-

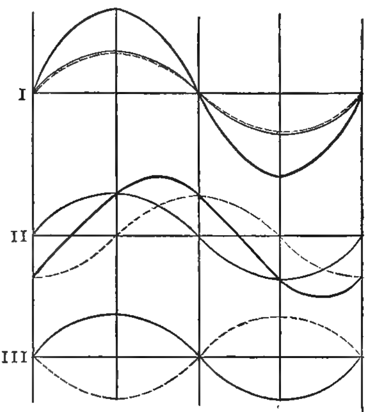

FIG. 47.

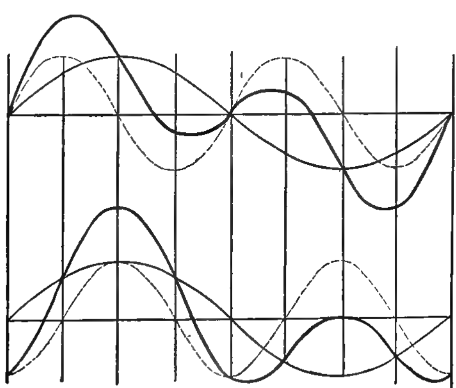

FIG. 48.

riods, and consequently the wave lengths, of the two systems are

as 1 : 2. It will be noticed that the resultant curve is no longer a simple sinusoid.

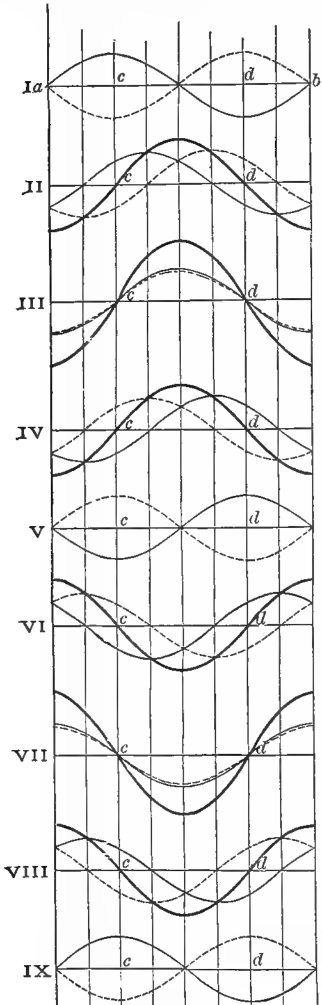

FIG. 49.

In the same way the resultant wave may be constructed for any number of wave systems having any relation of wave lengths, amplitudes, and phases. A very important case is that of two wave systems of the same period moving in opposite directions with the same velocity. In this case the two systems no longer maintain the same relative positions, and the resultant curve is not displaced along the axis, but continually changes form. In Fig. 49 let the full and dotted lines in I represent, at a given instant, the displacements due to the two waves respectively. The resultant is plainly the straight line *ab*, which indicates that at that instant there is no displacement of any particle. At an instant later by $\frac{1}{8}$ period, as shown in II, the wave represented by the full line has moved to the right $\frac{1}{8}$ wave length, while that represented by the dotted line has moved to the left the same distance. The heavy line indicates the corresponding displacements. In III, IV, V, etc., the conditions at instants $\frac{1}{4}$, $\frac{3}{8}$, $\frac{1}{2}$, etc., periods later are represented. A comparison of these figures will show that the particles at *c* and *d* are always at rest, that the particles between *c* and *d* all move in the same direction at the same time, and that particles on

the opposite sides of *c* or *d* are always moving in opposite directions. It follows that the resultant wave has no progressive motion. It is a *stationary wave*. Places where no motion occurs, such as *c* and *d*, are called *nodes*. The space between two nodes is an *internode* or *ventral segment*. The middle of a ventral segment, where the motion is greatest, is an *anti-node*. It will be seen later that all sounding bodies afford examples of stationary waves.

133. Reflection of Waves.—When a wave reaches the bounding surface between two media, one of three cases may occur:

(1) The particles of the second medium may have the same facility for movement as those of the first. The condition at the boundary will then be the same as that at any point previously traversed, and the wave will proceed as though the first medium were continuous.

(2) The particles of the second medium may move with less facility than those of the first. Then the condensed portion of a wave which reaches the boundary becomes more condensed in consequence of the restricted forward movement of the bounding particles, and the rarefied portion becomes more rarefied, because those particles are also restricted in their backward motion. The condensation and rarefaction are communicated backward from particle to particle of the first medium, and constitute a *reflected wave*. It will be seen that when the condensed portion of the wave, in which the particles have a forward movement, reaches the boundary, the effect is a greater condensation, that is, the same effect as would be produced by imparting a backward movement to the bounding particles if no wave previously existed. In the direct rarefied portion of the wave the movement of the particles is backward, and the effect, at the boundary, of a greater rarefaction is what would be produced by a forward movement of those particles. The effect in this case is, therefore, to reverse the motion of the particles. It is called *reflection with change of sign*.

(3) The particles of the second medium may move more freely than those of the first. In this case, when a wave in the first

medium reaches the boundary, the bounding particles, instead of stopping with a displacement such as they would reach in the interior of the medium, move to a greater distance, and this movement is communicated back from particle to particle as a reflected wave, in which the motion has the same sign as in the direct wave. It is *reflection without change of sign.* The two latter cases are extremely important in the study of the formation of stationary waves in sounding bodies.

Let us suppose a system of spherical waves departing from the point C (Fig. 50). Let mn be the intersection of one of the waves with the plane of the paper. Let AB be the trace of a plane smooth surface perpendicular to the plane of the paper, upon which the waves impinge. mo shows the position which the wave of which mn is a part would have occupied had it not been intercepted by the surface. From the last section it appears that reflection will take place as the wave mno strikes the various points of AB. In § 130 it was seen that any point of a wave may be considered as the centre of a wave system, and we may therefore take n', n'', etc., the points of intersection of the surface AB with the wave mn when it occupied the positions $m'n'$, $m''n''$, etc., as the centres of systems of spherical waves, the resultant of which would be the actual wave proceeding from AB. With n' as a centre describe a sphere tangent to mno at o. It is evident that this will represent the elementary spherical wave of which the centre is n' when the main wave is at mn. Describe similar spheres with n'', n''', etc., as centres. The surface np, which envelops and is tangent to all these spheres, represents the wave reflected from AB. If that part of the plane of the paper below AB be revolved about AB as an axis until it coincides with

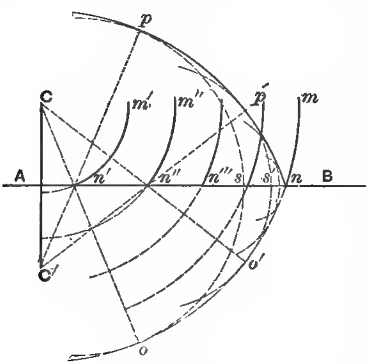

FIG. 50.

the paper above AB, *so* will coincide with sp, $s'o'$ with $s'p'$, etc., and hence *no* with *np*. But *no* is a circle with C as a centre; *np* is, therefore, a circle of which the centre is C', on a perpendicular to AB through C, and as far below AB as C is above. When, therefore, a wave is reflected at a plane surface, the centres of the incident and reflected waves are on the same line perpendicular to the reflecting surface, and at equal distances from the surface on opposite sides.

134. Theoretical Velocity of Sound.—The disturbance of the parts of any elastic medium which is propagating sound is assumed, in theoretical discussions, to take place in the line of direction of the propagation of the sound, and to be such that the type of the disturbance remains unaltered during its propagation. The velocity of propagation of such a disturbance may be investigated by the following method, due to Rankine.

Let us consider, as in § 129, a portion of the elastic medium in the form of an indefinitely long cylinder. If a disturbance be set up at any cross-section of this cylinder (Fig. 51), which consists of a displacement of the matter in that cross-section in the direction of the axis of the cylinder, it will, by hypothesis, be propagated in the direction of the axis with a constant velocity V, which is to be determined. If we consider any cross-section of the cylinder which is traversed by the disturbance, the matter which passes

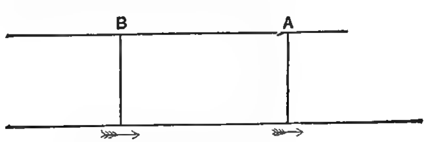

FIG. 51.

through it at any instant will have a velocity which may vary from zero to the maximum velocity of the vibrating matter, either positively when this velocity is in the direction of propagation of the disturbance, or negatively when it is opposite to it.

If we now conceive an imaginary cross-section A to move along the cylinder with the disturbance with the velocity V, the velocity

of the particles in it at any instant will be always the same. Let us call this velocity v_a. The velocity of the cross-section relative to the moving particles in it is then $V - v_a$. If we represent by d_a the density of the medium at the cross-section through which the velocity of the particles is v_a, which is the same for all positions of the moving cross-section, and if we assume that the area of the cross-section is unity, then the quantity of matter M which passes through the moving cross-section in unit time is $M = d_a(V - v_a)$.

If we conceive any other cross-section B to be moving with the disturbance in a similar manner, the same quantity of matter M will pass through it in unit time, since the two cross-sections move with the same velocity and the density of the matter between them remains the same. Hence we have $M = d_b(V - v_b)$, where d_b and v_b represent the quantities at the cross-section B corresponding to those at the cross-section A represented by d_a and v_a. Hence $d_a(V - v_a) = d_b(V - v_b)$. Since this equation is true whatever be the distance between the cross-sections, it is true for that position of the cross-section B for which $v_b = 0$, and for which $d_b = D$, the density of the medium in its undisturbed condition. Hence we have $M = DV$, $d_a(V - v_a) = DV$, and

$$\frac{v_a}{V} = \frac{d_a - D}{d_a}. \tag{59}$$

If the disturbance be small, the expression on the right is approximately the condensation per unit volume of the medium at the cross-section A, and the equation shows that the ratio of the velocity of the matter passing through the cross-section A to the velocity of propagation of the disturbance is equal to the condensation at that cross-section.

Now, to eliminate the unknown quantities v_a and d_a, we must find a new equation involving them. A quantity of matter M enters the region between the two moving cross-sections with the velocity v_a, and an equal quantity leaves the region with the velocity v_b. The difference of the momenta of the entering and outgoing quantities is $M(v_a - v_b)$. This difference can only be due to

the different pressures p_a and p_b on the moving cross-sections, since the interactions of the portion of matter between those cross-sections cannot change the momentum of that portion. Hence we have $M(v_a - v_b) = p_a - p_b$.

If we for convenience assume $v_b = 0$, we have $p_b = P$, the pressure in the medium in its undisturbed condition. If we further substitute for v_a its value, we obtain $MV = d_a \frac{p_a - P}{d_a - D}$. If the changes in pressure and density be small, the quantity $d_a \frac{p_a - P}{d_a - D}$ equals E, the modulus of elasticity of the medium. If we further substitute for M its value VD, we obtain finally

$$V^2 = \frac{E}{D} \quad \text{or} \quad V = \sqrt{\frac{E}{D}}. \tag{60}$$

135. Velocity of Sound in Air.—In air at constant temperature the elasticity is numerically equal to the pressure (§ 105). The compressions and rarefactions in a sound-wave occur so rapidly that during the passage of a wave there is no time for the transfer of heat, and the elasticity to be considered, therefore, is the elasticity when no heat enters or escapes (§ 213).

If the ratio of the two elasticities be represented by γ we have for the elasticity when no heat enters or escapes $E = \gamma P$, and the velocity of a sound-wave in air at zero temperature is given by $V = \sqrt{\frac{\gamma P}{D}}$. The coefficient γ equals 1.41. P is the pressure exerted by a column of mercury 76 centimetres high and with a cross-section of one square centimetre, or $76 \times 13.59 \times 981 = 1013373$ dynes per square centimetre. D equals 0.001293 gram at 0°, hence

$$V = \sqrt{\frac{1.41 \times 1013373}{0.001293}} = 33240, \text{ or } 332.4 \text{ metres per second.}$$

Since the density of air changes with the temperature, the velocity of sound must also change. If d_t represent the density at temperature t, and d_0 the density at zero, $d_t = \frac{d_0}{1 + \alpha t}$, from § 211.

The formula for velocity then becomes $V = \sqrt{\frac{\gamma P}{d_0}(1 + \alpha t)}$. This formula shows that the velocity at any temperature is the velocity at 0° multiplied by the square root of the factor of expansion.

136. Measurements of the Velocity of Sound.—The velocity of sound in air has been measured by observing the time required for the report of a gun to travel a known distance.

One of the best determinations was that made in Holland in 1822. Guns were fired alternately at two stations about nine miles apart. Observers at one station observed the time of seeing the flash and hearing the report from the other. The guns being fired alternately, and the sound travelling in opposite directions, the effect of wind was eliminated in the mean of the results at the two stations. It is possible, by causing the sound-wave to act upon diaphragms, to make it record its own time of departure and arrival, and by making use of some of the methods of estimating very small intervals of time the velocity of sound may be measured by experiments conducted within the limits of an ordinary building.

The velocity of sound in water was determined on Lake Geneva in 1826 by an experiment analogous to that by which the velocity in air was determined.

In § 144 and § 146 it is shown that the time of one vibration of any body vibrating longitudinally is the time required for a sound-wave to travel twice the distance between two nodes. The velocity may, therefore, be measured by determining the number of vibrations per second of the sound emitted, and measuring the distance between the nodes.

In an open organ-pipe, or a rod free at both ends, when the fundamental tone is sounded the sound travels twice the length of the rod or pipe during the time of one complete vibration. If rods of different materials be cut to such lengths that they all give the same fundamental tone when vibrating longitudinally, the ratio of their lengths will be that of the velocity of sound in them.

In Kundt's experiment, the end of a rod having a light disk attached is inserted in a glass tube containing a light powder strewn

over its inner surface. When the rod is made to vibrate longitudinally, the air-column in the tube, if of the proper length, is made to vibrate in unison with it. This agitates the powder and causes it to indicate the positions of the nodes in the vibrating air-column. The ratio of the velocity of sound in the solid to that in air is thus the ratio of the length of the rod to the distance between the nodes in the air-column.

CHAPTER II.

SOUNDS AND MUSIC.

COMPARISON OF SOUNDS.

137. Musical Tones and Noises.—The distinction between the impressions produced by musical tones and by noises is familiar to all. Physically, a musical *tone* is a sound the vibrations of which are regular and periodic. A *noise* is a sound the vibrations of which are very irregular. It may result from a confusion of musical tones, and is not always devoid of musical value. The sound produced by a block of wood dropped on the floor would not be called a musical tone, but if blocks of wood of proper shape and size be dropped upon the floor in succession, they will give the tones of the musical scale.

Musical tones may differ from one another in *pitch*, depending upon the frequency of the vibrations; in *loudness*, depending upon the amplitude of vibration; and in *quality*, depending upon the manner in which the vibration is executed. In regard to pitch, tones are distinguished as *high* or *low*, *acute* or *grave*. In regard to loudness, they are distinguished as *loud* or *soft*. The quality of musical tones enables us to distinguish the tones of different instruments even when sounding the same notes.

138. Methods of Determining the Number of Vibrations of a Musical Tone.—That the pitch of a tone depends upon the frequency of vibrations may be simply shown by holding the corner of

a card against the teeth of a revolving wheel. With a very slow motion the card snaps from tooth to tooth, making a succession of distinct taps, which, when the revolutions are sufficiently rapid, blend together and produce a continuous tone, the pitch of which rises and falls with the changes of speed. Savart made use of such a wheel to determine the number of vibrations corresponding to a tone of given pitch. After regulating the speed of rotation until the given pitch was reached, the number of revolutions per second was determined by a simple attachment; this number multiplied by the number of teeth in the wheel gave the number of vibrations per second.

The *siren* is an instrument for producing musical tones by puffs of air succeeding each other at short equal intervals. A circular disk having in it a series of equidistant holes arranged in a circle around its axis is supported so as to revolve parallel to and almost touching a metal plate in which is a similar series of holes. The plate forms one side of a small chamber, to which air is supplied from an organ bellows. If there be twenty holes in the disk, and if it be placed so that these holes correspond to those in the plate, air will escape through all of them. If the disk be turned through a small angle, the holes in the plate will be covered and the escape of air will cease. If the disk be turned still further, at one twentieth of a revolution from its first position, air will again escape, and if it rotate continuously, air will escape twenty times in a revolution. When the rotation is sufficiently rapid, a continuous tone is produced, the pitch of which rises as the speed increases. The siren may be used exactly as the toothed wheel to determine the number of vibrations corresponding to any tone.

By drilling the holes in the plate obliquely forward in the direction of rotation, and those in the disk obliquely backward, the escaping air will cause the disk to rotate, and the speed of rotation may be controlled by controlling the pressure of air in the chamber.

Sirens are sometimes made with several series of holes in the disk. These serve not only the purposes described above, but also to compare tones of which the vibration numbers have certain ratios.

The number of vibrations of a sounding body may sometimes be determined by attaching to it a light stylus which is made to trace a curve upon a smoked glass or cylinder. Instead of attaching a stylus to the sounding body directly, which is practicable only in a few cases, it may be attached to a membrane which is caused to vibrate by the sound-waves which the body generates. A membrane reproduces very faithfully all the characteristics of the sound-waves, and the curve traced by the stylus attached to it gives information, therefore, not only in regard to the number of vibrations, but to some extent in regard to their amplitude and form.

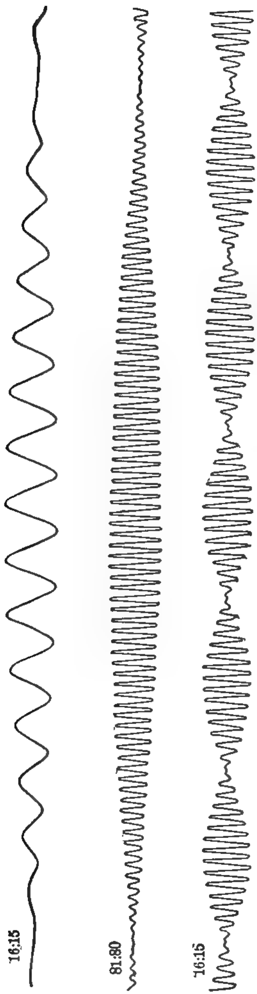

Fig. 52.

PHYSICAL THEORY OF MUSIC.

139. Concord and Discord.—When two or more tones are sounded together, if the effect be pleasing there is said to be *concord;* if harsh, *discord.* To understand the cause of discord, suppose two tones of nearly the same pitch to be sounded together. The resultant curve, constructed as in § 132, is like those in Fig. 52, which represent the resultants when the periods of the components have the ratio 81 : 80 and when they have the ratio 16 : 15. The figure indicates, what experiment verifies, that the resultant sound suffers periodic variations in intensity. When these variations occur at such intervals as to be readily distinguished, they are called *beats.* These beats occur more and more frequently as the numbers expressing the ratio of the

vibrations reduced to its lowest terms become smaller, until they are no longer distinguishable as separate beats, but appear as an unpleasant roughness in the sound. If the terms of the ratio become smaller still, the roughness diminishes, and when the ratio is $\frac{6}{5}$ the effect is no longer unpleasant. This, and ratios expressed by smaller numbers, as $\frac{5}{4}$, $\frac{5}{3}$, $\frac{4}{3}$, $\frac{3}{2}$, $\frac{2}{1}$, represent concordant combinations.

140. Major and Minor Triads.—Three tones of which the vibration numbers are as 4 : 5 : 6 form a concordant combination called the *major triad.* The ratio 10 : 12 : 15 represents another concordant combination called the *minor triad.* Fig. 53 shows the resultant curves for the two triads.

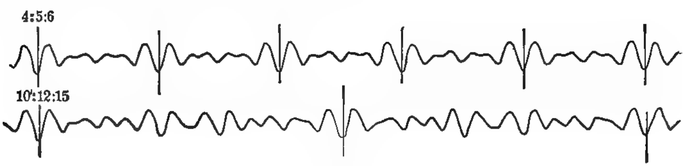

FIG. 53.

141. Intervals.—The *interval* between two tones is expressed by the ratio of their vibration numbers, using the larger as the numerator. Certain intervals have received names derived from the relative positions of the two tones in the musical scale, as described below. The interval $\frac{2}{1}$ is called an *octave;* $\frac{3}{2}$, a *fifth;* $\frac{4}{3}$, a *fourth;* $\frac{5}{4}$, a *major third;* $\frac{6}{5}$, a *minor third.*

142. Musical Scales.—A *musical scale* is a series of tones which have been chosen to meet the demands of musical composition. There are at present two principal scales in use, each consisting of seven notes, with their octaves, chosen with reference to their fitness to produce pleasing effects when used in combination. In one, called the *major scale,* the first, third, and fifth, the fourth, sixth, and eighth, and the fifth, seventh, and ninth tones, form major triads. In the other, called the *minor scale,* the same tones form minor triads. From this it is easy to deduce the following relations:

MAJOR SCALE.

								1′	2′
Tone Number.......... ..	1	2	3	4	5	6	7	8	9
Letter....................	C	D	E	F	G	A	B	C′	D′
Name.....................	do or ut	re	mi	fa	sol	la	si	ut	re
Number of vibrations.......	m	$\frac{9}{8}$m	$\frac{5}{4}$m	$\frac{4}{3}$m	$\frac{3}{2}$m	$\frac{5}{3}$m	$\frac{15}{8}$m	2m	$\frac{9}{4}$m
Intervals from tone to tone..	$\frac{9}{8}$	$\frac{10}{9}$	$\frac{16}{15}$	$\frac{9}{8}$	$\frac{10}{9}$	$\frac{9}{8}$	$\frac{16}{15}$		

MINOR SCALE.

Tone Number...............	1	2	3	4	5	6	7	8	9
Letter.....................	A	B	C	D	E	F	G	A′	B′
Name......................	la	si	ut	re	mi	fa	sol	la	si
Number of vibrations.......	m	$\frac{9}{8}$m	$\frac{6}{5}$m	$\frac{4}{3}$m	$\frac{3}{2}$m	$\frac{8}{5}$m	$\frac{9}{5}$m	2m	$\frac{9}{4}$m
Intervals from tone to tone..	$\frac{9}{8}$	$\frac{16}{15}$	$\frac{10}{9}$	$\frac{9}{8}$	$\frac{16}{15}$	$\frac{9}{8}$	$\frac{10}{9}$		

The derivation of the names of the intervals will now be apparent. For example, an interval of a third is the interval between any tone of the scale and the third one from it, counting the first as 1. If we consider the intervals from tone to tone, it is seen that the pitch does not rise by equal steps, but that there are three different intervals, $\frac{9}{8}$, $\frac{10}{9}$, and $\frac{16}{15}$. The first two are usually considered the same, and are called *whole tones.* The third is a *half-tone* or *semitone.*

It is desirable to be able to use any tone of a musical instrument as the first tone or *tonic* of a musical scale. To permit this, when the tones of the instrument are fixed, it is plain that extra tones, other than those of the simple scale, must be provided in order that the proper sequence of intervals may be maintained. Suppose the tonic to be transposed from C to D. The semitones should now come, in the major scale, between F and G, and C′ and D′, instead of between E and F, and B and C′. To accomplish this, a tone must be substituted for F and another for C′. These are called F sharp and C′ sharp respectively, and their vibration numbers are determined by multiplying the vibration numbers of the tones which they replace by $\frac{25}{24}$. The introduction of five such extra tones, making twelve in the octave, enables us to preserve the proper sequence of whole tones and semitones, whatever tone is

taken as the tonic. But if we consider that the whole tones are not all the same, and propose to preserve exactly all the intervals of the transposed scale, the problem becomes much more difficult, and can only be solved at the expense of too great complication in the instrument. Instead of attempting it, a system of tuning, called *temperament*, is used by which the twelve tones referred to above are made to serve for the several scales, so that while none are perfect, the imperfections are nowhere marked. The system of temperament usually employed, or at least aimed at, called the *even temperament*, divides the octave into twelve equal semitones, and each interval is therefore the twelfth root of 2. With instruments in which the tones are not fixed, like the violin for instance, the skilful performer may give them their exact value.

For convenience in the practice of music and in the construction of musical instruments, a *standard pitch* must be adopted. This pitch is usually determined by assigning a fixed vibration number to the tone above the middle C of the piano, represented by the letter A′. This number is about 440, but varies somewhat in different countries and at different times. In the instruments made by König for scientific purposes the vibration number 256 is assigned to the middle C. This has the advantage that the vibration numbers of the successive octaves of this tone are powers of 2.

CHAPTER III.

VIBRATIONS OF SOUNDING BODIES.

143. General Considerations.—The principles developed in § 133 apply directly in the study of the vibrations of sounding bodies. When any part of a body which is capable of acting as a sounding body is set in vibration, a wave is propagated through it to its boundaries, and is there reflected. The reflected wave, travelling away from the boundary, in conjunction with the direct wave going toward it, produces a stationary wave. These stationary waves are characteristic of the motion of all sounding bodies. Fixed points of a body often determine the position of nodes, and in all cases the length of the wave must have some relation to the dimensions of the body.

144. Organ Pipes.—A column of air, enclosed in a tube of suitable dimensions, may be made to vibrate and become a sounding body. Let us suppose a tube closed at one end and open at the other. If the air particles at the open end be suddenly moved inward, a pulse travels to the closed end, and is there reflected with change of sign (§ 133). It returns to the open end and is again reflected, this time without change of sign, because there is greater freedom of motion without than within the tube. As it starts again toward the closed end, the air particles that compose it move outward instead of inward. If they now receive an independent impulse outward, the two effects are added and a greater disturbance results. So, by properly timing small impulses at the open end of the tube, the air in it may be made to vibrate strongly.

If a continuous vibration be maintained at the open end of the tube, waves follow each other up the tube, are reflected with

change of sign at the closed end, and returning, are reflected without change of sign at the open end. Any given wave *a*, therefore, starts up the tube the second time with its phase changed by half a period. The direct wave that starts up the tube at the same instant must be in the same phase as the reflected wave, and it therefore differs in phase half a period from the direct wave *a*. In other words, any wave returning to the mouthpiece must find the vibrations there opposite in phase to those which existed when it left. This is possible only when the vibrating body makes, during the time the wave is going up the tube and back, 1, 3, 5, or some odd number of half-vibrations. By constructing the curves representing the stationary wave resulting from the superposition of the two systems of vibrations, it will be seen that there is always a node at the closed end of the tube and an anti-node at the mouth. When there is 1 half-vibration while the wave travels up and back, the length of the tube is $\frac{1}{4}$ the wave length; when there are 3 half-vibrations in the same time, the length of the tube is $\frac{3}{4}$ the wave length, and there is a node at one third the length of the tube from the mouth.

If the tube be open at both ends, reflection without change of sign takes place in both cases, and the reflected wave starts up the tube the second time in the same phase as at first. The vibrations must therefore be so timed that 1, 2, 3, 4, or some whole number of complete vibrations are performed while the wave travels up the tube and back. A construction of the curve representing the stationary wave in this case will show, for the smallest number of vibrations, a node in the middle of the tube and an anti-node at each end. The length of the tube is therefore $\frac{1}{2}$ the wave length for this rate of vibration. The vibration numbers of the several tones produced by an open tube are evidently in the ratio of the series of whole numbers 1, 2, 3, 4, etc., while for the closed tube only those tones can be produced of which the vibration numbers are in the ratio of the series of odd numbers 1, 3, 5, etc. It is evident also that the lowest tone of the closed tube is an octave lower than that of the open tube of the same length.

This lowest tone of the tube is called the *fundamental*, and the others are called *overtones*, or *harmonics*. These simple relations between the length of the tube and length of the wave are only realized when the tubes are so narrow that the air particles lying in a plane cross-section are all actuated by the same movement. This is never the case at the open end of the tube, and the distance from this end to the first node is, therefore, always less than a quarter wave length.

145. Modes of Exciting Vibrations in Tubes.—If a tuning-fork be held in front of the open mouth of a tube of proper length, the sound of the fork is strongly reinforced by the vibration of the air in the tube. If we merely blow across the open end of a tube, the agitation of the air may, by the reaction of the returning reflected pulses, be made to assume a regular vibration of the proper rate and the column made to sound. In organ pipes a mouthpiece of the form shown in Fig. 54 is often employed. The thin sheet of air projected against the thin edge is thrown into vibration. Those elements of this vibration which correspond in frequency with the pitch of the pipe are strongly reinforced by the action of the stationary wave set up in the pipe, and hence the tone proper to the pipe is produced.

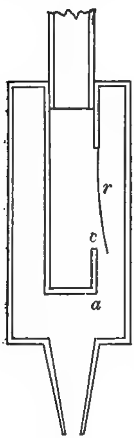

FIG. 54*a*.

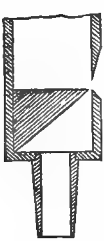
FIG. 54.

Sometimes *reeds* are used, as shown in Fig. 54*a*. The air escaping from the chamber *a* through the passage *c* causes the reed *r* to vibrate. This alternately closes and opens the passage, and so throws into vibration the air in the pipe. If the reed be stiff, and have a determined period of vibration of its own, it must be tuned to suit the period of the air-column which it is intended to set in vibration. If the reed be very flexible, it will accommodate itself to the rate of vibration of the air-column, and may then serve to produce various tones, as in the clarionet.

In instruments like the cornet and bugle the lips of the player

act as a reed, and the player may at will produce many of the different overtones. In that way melodies may be played without the use of keys or other devices for changing the length of the air-column.

Vibrations may be excited in a tube by placing a gas flame at the proper point in it. The flame thus employed is called a *singing flame.* The organ of the voice is a kind of reed pipe in which little folds of membrane, called vocal chords, serve as reeds which can be tuned to different pitches by muscular effort, and the cavity of the mouth and larynx serves as a pipe in which the mass of air may also be changed at will, in form and volume.

146. Longitudinal Vibrations of Rods.—A rod free at both ends vibrates as the column of air in an open tube. Any displacement produced at one end is transmitted with the velocity of sound in the material to the other end, is there reflected without change of sign, and returns to the starting-point to be reflected again exactly as in the open tube. The fundamental tone corresponds to a stationary wave having a node at the centre of the rod.

147. Longitudinal Vibrations of Cords.—Cords fixed at both ends may be made to vibrate by rubbing them lengthwise. Here reflection with change of sign takes place at both ends, which brings the wave as it leaves the starting-point the second time to the same phase as when it first left it, and there must be, therefore, as in the open tube, 1, 2, 3, 4, etc., vibrations while the wave travels twice the length of the cord. The velocity of transmission of a longitudinal displacement in a wire depends upon the elasticity and density of the material only. The velocity and the rate of vibration are, therefore, nearly independent of the stretching force.

148. Transverse Vibrations of Cords.—If a transverse vibration be given to a point upon a wire fastened at both ends, everything relating to the reflection of the wave motion and the formation of stationary waves is the same as for longitudinal displacements. The velocity of transmission, and consequently the frequency of the vibrations, are, however, very different. They depend on the stretching force or tension and on the mass of the cord per unit

length. The number of vibrations is inversely as the length of the cord, directly as the square root of the tension, and inversely as the square root of the mass per unit length.

149. Transverse Vibrations of Rods, Plates, etc.—The vibrations of rods, plates, and bells are all cases of stationary waves resulting from systems of waves travelling in opposite directions. Subdivision into segments occurs, but in these cases the relations of the various overtones are not so simple as in the cases before considered. For a rod fixed at one end, sounding its fundamental tone, there is a node at the fixed end only. For the first overtone there is a second node near the free end of the rod, and the number of vibrations is a little more than six times the number for the fundamental.

A rod free at both ends has two nodes when sounding its fundamental, as shown in Fig. 55. The distance of these nodes from the ends is about $\frac{2}{9}$ the length of the rod. If the rod be bent, the nodes approach the centre until, when it has assumed the U form like a tuning-fork, the two nodes are very near the centre. This will be understood from Fig. 56.

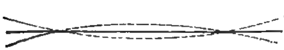

FIG. 55.

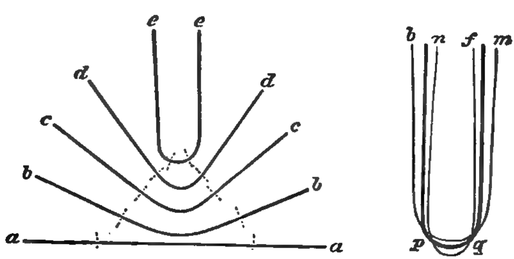

FIG. 56.

The nodal lines on plates may be shown by fixing the plate in a horizontal position and sprinkling sand over its surface. When the plate is made to vibrate, the sand gathers at the nodes and marks their position. The figures thus formed are known as *Chladni's figures.*

150. Resonance.—If several pendulums be suspended from the

same support, and one of them be made to vibrate, any others which have the same period of vibration will soon be found in motion, while those which have a different period will show no signs of disturbance. The vibration of the first pendulum produces a slight movement of the support, which is communicated alike to all the other pendulums. Each movement may be considered as a slight impulse, which imparts to each pendulum a very small vibratory motion. For those pendulums having the same period as the one in vibration, these impulses come just in time to increase the motion already produced, and so, after a time, produce a sensible motion; while for those pendulums having a different period the vibration at first imparted will not keep time with the impulses, and these will therefore as often tend to destroy as to increase the motion. It is important to note that the pendulum imparting the motion loses all it imparts. This is not only true of pendulums, but of all vibrating bodies. Two strings stretched from the same support and tuned to unison will both vibrate when either one is caused to sound. A tuning-fork suitably mounted on a sounding-box will communicate its vibrations to another tuned to exact unison even when they are thirty or forty feet apart and only air intervenes. In this case it is the sound-wave generated by the first fork which excites the second fork, and in so doing the wave loses a part of its own motion, so that beyond the second fork, on the line joining the two, the sound will be less intense than at the same distance in other directions. Such communication of vibrations is called *resonance*.

Air-columns of suitable dimensions will vibrate in sympathy with other sounding bodies. If water be gradually poured into a deep jar, over the mouth of which is a vibrating tuning-fork, there will be found in general a certain length of the air-column for which the tone of the fork is strongly reinforced. From the theory of organ pipes, it is plain that this length corresponds approximately to a quarter wave length for that tone. In this case, also, when the strongest reinforcement occurs, the sound of the fork will rapidly die away. The sounding-boxes on which the tun-

ing-forks made by König are mounted are of such dimensions that the enclosed body of air will vibrate in unison with the fork, but they are purposely made not quite of the dimensions for the best resonance, in order that the forks may not too quickly be brought to rest.

A membrane or a disk, fastened by its edges, may respond to and reproduce more or less faithfully a great variety of sounds. Hence such disks, or *diaphragms*, are used in instruments like the telephone and phonograph, designed to reproduce the sounds of the voice. The *phonograph* consists of a mouthpiece and disk similar to that used in the telephone, but the disk has fastened to its centre, on the side opposite the mouthpiece, a short stiff stylus, which serves to record the vibrations of the disk upon a sheet of tinfoil or wax moved along beneath it. The wax is in the form of a cylinder mounted on an axle moved by a screw working in a fixed nut, so that when the cylinder revolves it has also an endwise motion, such that a fixed point would follow a spiral track on its surface. To use the instrument, the disk is placed in position with the stylus attached and slightly indenting the wax. The cylinder is revolved while sounds are produced in front of the disk. The disk vibrates, causing the stylus to indent the wax more or less deeply, so leaving a permanent record. If now the cylinder be turned back to the starting-point and then turned forward, causing the stylus to go over again the same path, the indentations previously made in the wax now cause the stylus, and consequently the disk, to vibrate and reproduce the sound that produced the record.

The sounding-boards of the various stringed instruments are in effect thin disks, and afford examples of the reinforcement of vibrations of widely different pitch and quality by the same body. The strings of an instrument are of themselves insufficient to communicate to the air their vibrations, and it is only through the sounding-board that the vibrations of the string can give rise to audible sounds. The quality of stringed instruments, therefore, depends largely upon the character of the sounding-board.

The tympanum of the ear furnishes another example of the facility with which membranes respond to a great variety of sounds.

CHAPTER IV.

ANALYSIS OF SOUNDS AND SOUND SENSATIONS.

151. Quality.—As has already been stated, the tones of different instruments, although of the same pitch and intensity, are distinguished by their *quality*. It was also stated that the quality of a tone depends upon the manner in which the vibration is executed. The meaning of this statement can best be understood by

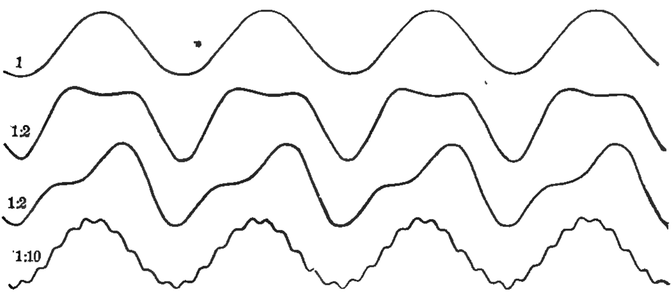

FIG. 57.

considering the curves which represent the vibrations. In Fig. 57 are given several forms of vibration curves of the same period.

Every continuous musical tone must result from a periodic vibration, that is, a vibration which, however complicated it may be, repeats itself at least as frequently as do the vibrations of the lowest audible tone. According to Fourier's theorem (§ 21), every periodic vibration is resolvable into simple harmonic vibrations having commensurable periods. It has been seen that all sounding bodies may subdivide into segments, and produce a series of tones of which the vibration periods generally bear a simple rela-

tion to one another. These may be produced simultaneously by the same body, and so give rise to complex tones, the character of which will vary with the nature and intensity of the simple tones produced. It has been held that the quality of a complex tone is not affected by change of phase of the component simple tones relative to each other. Some experiments by König seem to indicate, however, that the quality does change when there is merely change of phase.

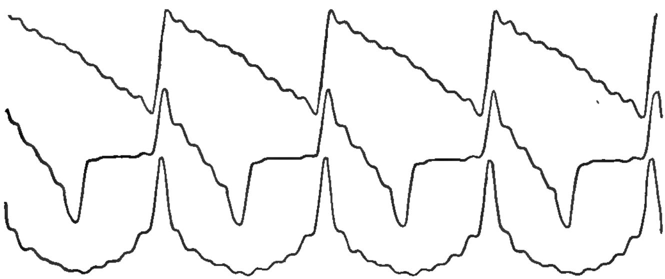

Fig. 58.

In Fig. 58 are shown three curves, each representing a fundamental accompanied by the harmonics up to the tenth. The curves differ only in the different phases of the components relative to each other.

Fig. 59 shows similar curves produced by a fundamental accompanied by the odd harmonics.

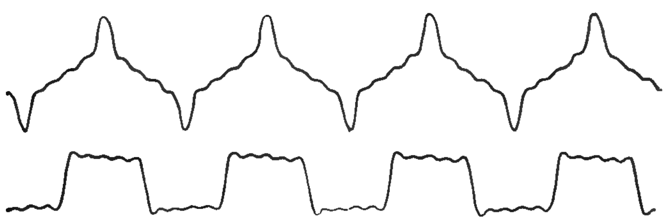

Fig. 59.

152. Resonators for the Study of Complex Tones.—An apparatus devised by Helmholtz serves to analyze complex tones and indicate the simple tones of which they are composed. It consists of a series

of hollow spheres or cylinders, called *resonators*, which are tuned to certain tones. If a tube lead from the resonator to the ear and a

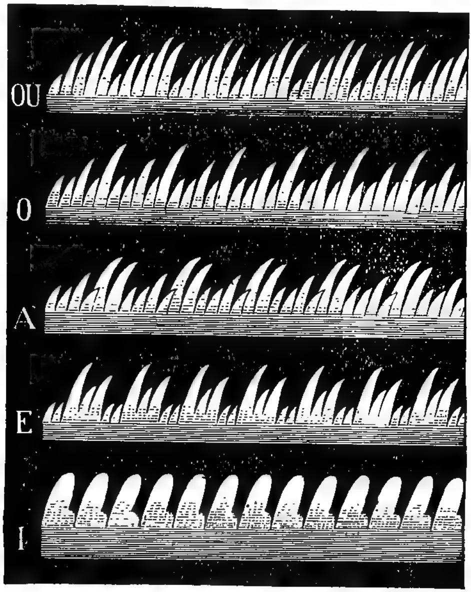

FIG. 60.

sound be produced, one of the components of which is the tone to which the resonator is tuned, the mass of air in it will be set in vibration, and that tone will be clearly heard ; or, if the resonator be connected by a rubber tube to a manometric capsule (§ 128), the gas flame connected with the capsule will be disturbed whenever the tone to which the resonator is tuned is produced in the vicinity, either by itself or as a component of a complex tone. By trying the resonators of a series, one after another, the several components of a complex tone may be detected and its composition demonstrated.

153. Vowel Sounds.—Helmholtz has shown that the differences between the vowel sounds are only differences of quality. That the vowel sounds correspond to distinct forms of vibration is well shown by means of the manometric flame. By connecting a mouthpiece

to the rear of the capsule, and singing into it the different vowel sounds, the flame images assume distinct forms for each. Some of these forms are shown in Fig. 60.

154. Optical Method of Studying Vibrations. — The vibratory motion of sounding bodies may sometimes be studied to advantage

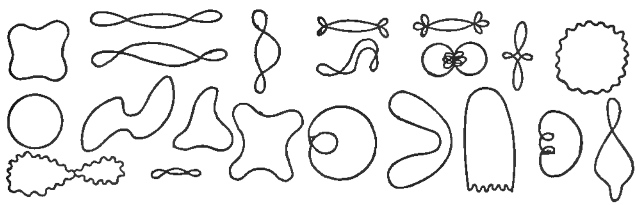

FIG. 61.

by observing the lines traced by luminous points upon the vibrating body or by observing the movement of a beam of light reflected from a mirror attached to the body.

Young studied the vibrations of strings by placing the string where a thin sheet of light would fall across it, so as to illuminate a single point. When the string was caused to vibrate, the path of the point appeared as a continuous line, in consequence of the persistence of vision. Some of the results which he obtained are given in Fig. 61, taken from Tyndall on Sound.

The most interesting application of this method was made by Lissajous to illustrate the composition of vibratory motions at right angles to each other. If a beam of light be reflected to a screen from a mirror attached to a tuning-fork, when the tuning-fork vibrates the spot on the screen will describe a simple harmonic motion and will appear as a straight line of light. If the beam, instead of being reflected to a screen, fall upon a mirror attached to a second fork, mounted so as to vibrate in a plane at right angles to the first, the spot of light will, when both forks vibrate, be actuated by two simple harmonic motions at right angles to each other, and the resultant path will appear as a curve more or less complicated, depending upon the relation of the two forks to each other as to both period and phase (§ 21). Fig. 62 shows some of

the simpler forms of these curves. The figures of the upper line are those produced by two forks in unison; those of the second

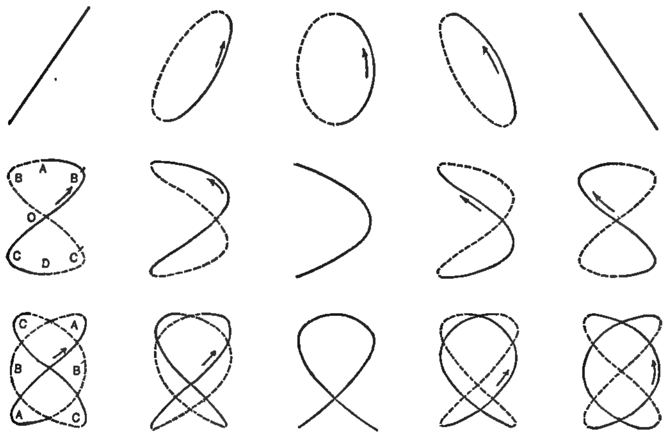

FIG. 62.

line by two forks of which the vibration numbers are as 2 : 1; those of the lower line by two forks of which the vibration numbers are as 3 : 2.

155. Beats.—It has already been explained (§ 139) that, when two tones of nearly the same pitch are sounded together, variations of intensity, called *beats*, are heard. Helmholtz's theory of the perception of beats was, that, of the little fibres in the ear which are tuned so as to vibrate with the various tones, those which are nearly in unison affect one another so as to increase and diminish one another's motions, and hence that no beats could be perceived unless the tones were nearly in unison. Beats are, however, heard when a tone and its octave are not quite in tune, and, in general, a tone making n vibrations produces m beats when sounded with a tone making $2n \pm m$, $3n \pm m$, etc., vibrations. This was explained in accordance with Helmholtz's theory, by assuming that one of the harmonics of the lower tone, which is nearly in unison with the upper, causes the beats, or, in cases where this is inadmissible, that they are caused by the lower tone in conjunction with a resultant

tone (§ 156). An exhaustive research by König, however, has demonstrated that beats are perceived when neither of the above suppositions is admissible. Figs. 63 and 64 show that the resultant

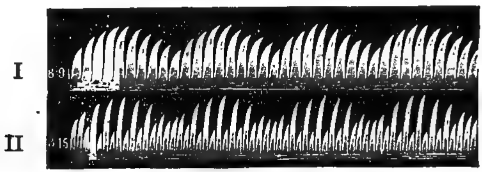

Fig. 63.

vibrations are affected by changes of amplitude similar to, though less in extent than, the changes which occur when the tones are nearly in unison. In Fig. 63 I represents a flame image obtained

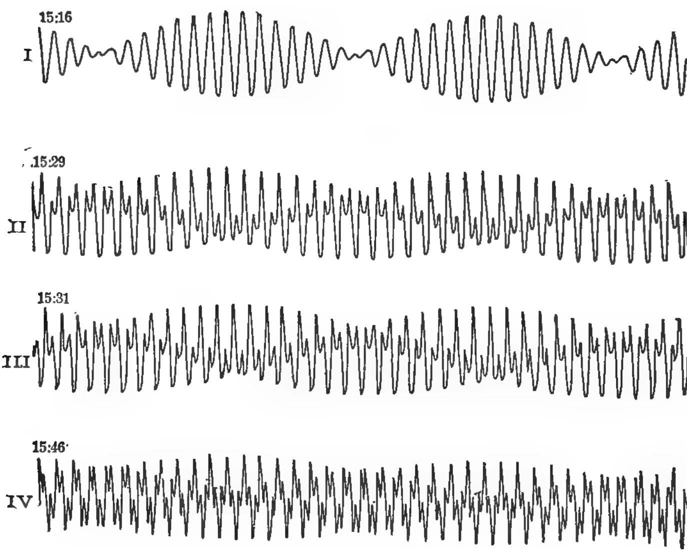

Fig. 64.

when two tones making n and $n \pm m$ vibrations, respectively, are produced together, and II represents the image when the number

of vibrations are n and $2n \pm m$. Fig. 64 shows traces obtained mechanically. In I the numbers of the component vibrations were n and $n + m$, in II and III n and $2n \pm m$, and in IV n and $3n + m$. In all these cases a variation of amplitude occurs during the same intervals, and it seems reasonable to suppose that those variations of amplitude should cause variations in intensity in the sound perceived.

Cross has shown that the beating of two tones is perfectly well perceived when the tones themselves are heard separately by the two ears; one tone being heard directly by one ear, while the other, produced in a distant room, is heard by the other ear by means of a telephone. Beats are also perceived when tones are produced at a distance from each other and from the listener, who hears them by means of separate telephones through separate lines. In this case there is no possibility of the formation of a resultant wave, or of any combination of the two sounds in the ear.

156. Resultant Tones.—*Resultant tones* are produced by combinations of two tones. Those most generally recognized have a vibration number equal to the sum or difference of the vibration numbers of their primaries. For instance, ut_6, making 2048 vibrations, and re_6, making 2304 vibrations, when sounded together give ut_3, making 256 vibrations. These tones are only heard well when the primaries are loud, and it requires an effort of the attention and some experience to hear them at all. Summation tones are more difficult to recognize than difference tones, nevertheless they have an influence in determining the general effect produced when musical tones are sounded together. Other resultant tones may be heard under favorable conditions. As described above, two tones making n and $n + m$ vibrations respectively, when m is considerably less than n, give a resultant tone making m vibrations; but a tone making n vibrations in combination with one making $2n + m$, $3n + m$, or $xn + m$ vibrations, gives the same resultant. This has sometimes been explained by assuming that intermediate resultants are produced, which, with one of the primaries, produce resultants of a higher order. In the case of the two tones making n and

$3n + m$ vibrations, for instance, the first difference tone would make $2n + m$ vibrations. This tone and the one making n vibrations would give the tone making $n + m$ vibrations; this tone, in turn, and the one making n vibrations would give the tone making m vibrations. This last tone is the one which is heard most plainly, and it seems difficult to admit that it can be the resultant of tones which are only heard very feebly, and often not at all. In Fig. 64 are represented the resultant curves produced in several of these cases. The first curve corresponds to two tones of which the vibration numbers are as 15 : 16. It shows the periodic increase and decrease in amplitude, occurring once every 15 vibrations, which, if not too frequent, give rise to beats (§ 139). If the pitch of the primaries be raised, preserving the relation 15 : 16, the beats become more frequent, and finally a distinct tone is heard, the vibration number of which corresponds to the number of beats that should exist. It was for a long time considered that the resultant tone was merely the rapid recurrence of beats. Helmholtz has shown by a mathematical investigation that a distinct wave making m vibrations will result from the coexistence of two waves making n and $n + m$ vibrations, and he believes that mere alternations of intensity, such as constitute beats, occurring ever so rapidly cannot produce a tone.

In II and III (Fig. 64) are the curves resulting from two tones, the intervals between which are respectively $15 : 29 (= 2 \times 15 - 1)$ and $15 : 31 (= 2 \times 15 + 1)$. Running through these may be seen a periodic change corresponding exactly in period to that shown in I. The same is true also of the curve in IV, which is the resultant for two tones the interval between which is $15 : 46 (= 3 \times 15 + 1)$. In all these cases, as has been already said (§ 155), if the pitch of the components be not too high, one beat is heard for every 15 vibrations of the lower component. Fig. 63 shows the flame images for the intervals $n : n + m$ and $n : 2n + m$. The varying amplitudes resulting in m beats per second are very evident in both. In all these cases, also, as the pitch of the components rises the beats become more frequent, and finally a resultant tone is heard, having, as already stated, one vibration for every 15 vibrations of the lower

component. In Fig. 65 are shown two resultant curves having

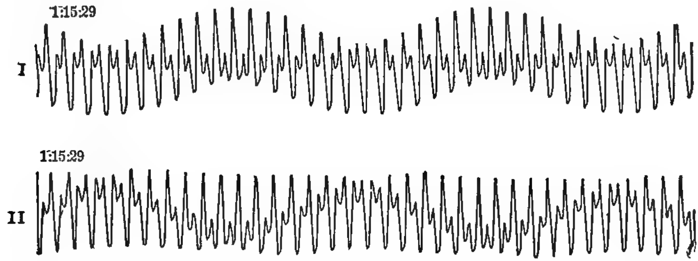

FIG. 65.

three components of which the vibration numbers are as 1 : 15 : 29. In I the three components all start in the same phase. In II, when 15 and 29 are in the same phase, I is in the opposite phase.

HEAT.

CHAPTER I.

MEASUREMENT OF HEAT.

157. General Effects of Heat.—Bodies are warmed, or their temperature is raised, by heat. The sense of touch is often sufficient to show difference in temperature; but the true criterion is the transfer of heat from the hotter to the colder body when the two bodies are brought in contact, and no work is done by one upon the other. This transfer is known by some of the effects described below.

Bodies, in general, expand when heated. Experiments show that different substances expand differently for the same rise of temperature. Gases, in general, expand more than liquids, and liquids more than solids. Expansion, however, does not universally accompany rise of temperature. A few substances contract when heated.

Heat changes the state of aggregation of bodies, always in such a way as to admit of greater freedom of motion among the molecules. The melting of ice and the conversion of water into steam are familiar examples.

Heat breaks up chemical compounds. The compounds of sodium, potassium, lithium, and other metals, give to the flame of a Bunsen lamp the characteristic colors of the vapors of the metals

which they contain. This fact shows that the heat separates the metals from their combinations.

When the junction of two dissimilar metals in a conducting circuit is heated, electric currents are produced.

Heat performs mechanical work. For example, the heat produced in the furnace of a steam-boiler may be used to drive an engine.

158. Production of Heat.—Heat is produced by various processes, some of which are the reverse of the operations just mentioned as the effects of heat. As examples of such reverse operations may be mentioned, the production of heat by the compression of a body which expands when heated; the production of heat during a change in the state of aggregation of a body, when the freedom of motion among the molecules is diminished; the production of heat during chemical combination; and the production of heat when an electric current passes through a junction of two dissimilar metals in an opposite direction to that of the current which is set up when the junction is heated.

Heat is produced in general in any process involving the expenditure of mechanical energy. The heat produced in such processes cannot be used to restore the whole of the original mechanical energy. The production of heat by friction is the best example of these processes.

Further, an electric current, in a homogeneous conductor, generates heat at every point in it, while if every point in the conductor be equally heated no current will be set up.

These cases are examples of the production of heat by non-reversible processes.

159. Nature of Heat.—Heat was formerly considered to be a substance which passed from one body to another, lowering the temperature of the one and raising that of the other, which combined with solids to form liquids, and with liquids to form gases or vapors. But the most delicate balances fail to show any change of weight when heat passes from one body to another. Rumford was able to raise a considerable quantity of water to the

boiling-point by the friction of a blunt boring-tool within the bore of a cannon. He showed that the heat manifested in this experiment could not have come from any of the bodies present, and also that heat would continue to be developed as long as the borer continued to revolve, or that the supply of heat was practically inexhaustible. The heat, therefore, must have been generated by the friction.

That ice is not melted by the combination with it of a heat substance was shown early in the present century by Davy. He caused ice to melt by friction of one piece upon another in a vacuum, the experiment being performed in a room where the temperature was below the melting-point of ice. There was no source from which heat could be drawn. The ice must, therefore, have been melted by the friction.

Rumford was convinced that the heat obtained in his experiment was only transformed mechanical energy; but to demonstrate this it was necessary to prove that the quantity of heat produced was always proportional to the quantity of mechanical work done. This was done in the most complete manner by Joule in a series of experiments extending from 1842 to 1849. He showed that, however the heat was produced by mechanical means, whether by the agitation of water by a paddle-wheel, the agitation of mercury, or the friction of iron plates upon each other, the same expenditure of mechanical energy always developed the same quantity of heat. Joule also proved the perfect equivalence of heat and electrical energy.

These experiments prove that *heat is a form of energy.* Consistent explanations of most if not all of the phenomena of heat may be given if we assume that the molecules of bodies, and the atoms constituting the molecules, are in constant motion, that the temperature of a body varies with the mean kinetic energy of an atom, and that the heat in a body is the sum of the kinetic energies of its atoms.

THERMOMETRY.

160. Temperature.—Two bodies are said to be at the same *temperature* when, if they be brought into each other's presence, no heat is transferred from one to the other. A body is at a *higher temperature* than the other bodies around it when it gives up heat to them. The fact that it gives up heat may be shown by its change in volume. A body is at a *lower temperature* when it receives heat from surrounding bodies. It is understood, of course, in what is said above, that one body has no action upon the other, or that no work is done by one body upon the other.

161. Thermometers.—Experiments show that, in general, bodies expand, and their temperature rises progressively, with the application of heat. An instrument may be constructed which will show at any instant the volume of a body selected for the purpose. If the volume increase, we know that the temperature rises; if the volume remain constant or diminish, we know that the temperature remains stationary or falls. Such an instrument is called a *thermometer.*

The thermometer most in use consists of a glass bulb with a fine tube attached. The bulb and part of the tube contain mercury. In order that the thermometers of different makers may give similar readings, it is necessary to adopt two standard temperatures which can be easily and certainly reproduced. The temperatures adopted are the melting-point of ice, and the temperature of steam from boiling water, under a pressure equal to that of a column of mercury 760 millimetres high at Paris. After the instrument has been filled with mercury, it is plunged in melting ice, from which the water is allowed to drain away, and a mark is made upon the stem opposite the end of the mercury column. It is then placed in a vessel in which water is boiled, so constructed that the steam rises through a tube surrounding the thermometer, and then descends by an annular space between that tube and an outer one, and escapes at the bottom. The thermometer does not touch the water, but is entirely surrounded by steam. The point

reached by the end of the mercury column is marked on the stem, as before. The space between these two marks is then divided into a number of equal parts.

While all makers of thermometers have adopted the same standard temperatures for the fixed points of the scale, they differ as to the number of divisions between these points. The thermometers used for scientific purposes, and in general use in France, have the space between the fixed points divided into a hundred equal parts or *degrees*. The melting-point of ice is marked 0°, and the boiling-point 100°. This scale is called the *Centigrade* or *Celsius* scale.

The *Réaumur* scale, in use in Germany, has eighty degrees between the melting- and boiling-points, and the boiling-point is marked 80°.

The *Fahrenheit* scale, in general use in England and America, has a hundred and eighty degrees between the melting- and boiling-points. The former is marked 32°, and the latter 212°.

The divisions in all these cases are extended below the zero point, and are numbered from zero downward. Temperatures below zero must, therefore, be read and treated as negative quantities.

A few points in the process of construction of a thermometer deserve notice. It is found that glass, after it has been heated to a high temperature, and again cooled, does not for some time return to its original volume. The bulb of a thermometer must be heated in the process of filling with mercury, and it will not return to its normal volume for some months. The construction of the scale should not be proceeded with until the reservoir has ceased to contract. For the same reason, if the thermometer be used for high temperatures, even the temperature of boiling water, time must be given for the reservoir to return to its original volume before it is used for the measurement of low temperatures.

It is essential that the diameter of the tube should be nearly uniform throughout, and that the divisions of the scale should represent equal capacities in the tube. To test the tube a thread of mercury about 50 millimetres long is introduced, and its length is

measured in different parts of the tube. If the length vary by more than a millimetre, the tube should be rejected. If the tube be found to be suitable, a bulb is attached, mercury is introduced, and the tube sealed after the mercury has been heated to expel the air. When it is ready for graduation, the fixed points are determined; then a thread of mercury having a length equal to about ten degrees of the scale is detached from the column, and its length measured in all parts of the tube. By reference to these measurements, the tube is so graduated that the divisions represent parts of equal capacity, and are not necessarily of equal length.

If such a thermometer indicate a temperature of 10°, this means that the thermometer is in such a thermal condition that the volume of the mercury has increased from zero one tenth of its total expansion from zero to 100°. There is no reason for supposing that this represents the same proportional rise of temperature. If a thermometer be constructed in the manner described, using some liquid other than mercury, it will not in general indicate the same temperature as the mercurial thermometer, except at the two standard points. It is plain, therefore, that a given fraction of the expansion of a liquid from zero to 100° cannot be taken as representing the same fraction of the rise of temperature.

162. Air-thermometer.—If a gas be heated, and its volume kept constant, its pressure increases. For all the so-called permanent gases—that is, those which are liquefied only with great difficulty—the amount of increase in pressure for the same increase of temperature is found to be almost exactly the same. This fact is a reason for supposing that the increase of pressure is proportional to the increase of temperature. There are theoretical reasons, as will be seen later, for the same supposition.

An instrument constructed to take advantage of this increase in pressure to measure temperature is called an *air-thermometer*. A bulb so arranged that it may be placed in the medium of which the temperature is to be determined, is filled with air or some other gas, and means are provided for maintaining the volume of the gas constant, and measuring its pressure. For the reasons given above,

the air-thermometer is taken as the standard instrument for scientific purposes. Its use, however, involves several careful observations and tedious computations. It is therefore mainly employed as a standard with which to compare other instruments. If we make such a comparison, and construct a table of corrections, we may reduce the readings of any thermometer to the corresponding readings of the air-thermometer.

163. Limits in the Range of the Mercurial Thermometer.—The range of temperature for which the mercurial thermometer may be employed is limited by the freezing of the mercury on the one hand, and its boiling on the other. For temperatures below the freezing-point of mercury alcohol thermometers may be used. For the measurement of high temperatures several different methods have been employed. One depends upon the expansion of a bar of platinum, another upon the variation in the electric resistance of platinum wire, another upon the strength of the electric current generated by a thermo-electric pair, another on the density of mercury vapor.

The special devices used in applying these methods need not be considered here.

CALORIMETRY.

164. Unit of Heat.—It is evident that more heat is required to raise the temperature of a large quantity of a substance through a given number of degrees than to raise the temperature of a small quantity of the same substance through the same number of degrees. It is further evident that the successive repetition of any operation by which heat is produced will generate more heat than a single operation. Heat is therefore a quantity the magnitude of which may be expressed in terms of some unit. The unit of heat generally adopted is the heat required to raise the temperature of one kilogram of water from zero to one degree. It is called a *calorie.*

It is sometimes convenient to employ a smaller unit, namely, the quantity of heat necessary to raise one gram of water from zero to one degree. This unit is designated as the *lesser calorie* or the

gram-degree. It is one one-thousandth of the larger unit. It may, therefore, be called a *millicalorie.*

The fact that heat is energy enables us to employ still another unit. It is that quantity of heat which is equivalent to an erg. This unit is called the *mechanical unit* of heat. According to the determination of Griffiths, a calorie contains about 41,982,000,000 mechanical units.

165. Heat required to raise the Temperature of a Mass of Water.—It is evident that to raise the temperature of m kilograms of water from zero to one degree will require m calories. If the temperature of the same quantity of water fall from one degree to zero, the same quantity of heat is given to surrounding bodies.

Experiment shows, that if the same quantity of water be raised to different temperatures, quantities of heat nearly proportional to the rise in temperature will be required: hence, to raise the temperature of m kilograms of water from zero to t degrees requires mt calories very nearly. This is shown by mixing water at a lower temperature with water at a higher temperature. The temperature of the mixture will be almost exactly the mean of the two. Regnault, who tried this experiment with the greatest care, found the temperature of the mixture a little higher than the mean, and concluded that the quantity of heat required to raise the temperature of a kilogram of water one degree increases slightly with the temperature; that is, to raise the temperature of a kilogram of water from twenty to twenty-one degrees, requires a little more heat than to raise the temperature of the same quantity of water from zero to one degree.

Rowland found, by mixing water at various temperatures, and also by measuring the energy required to raise the temperature of water by agitation by a paddle-wheel, that, when the air thermometer is taken as a standard, the quantity of heat necessary to raise the temperature of a given quantity of water one degree diminishes slightly from zero to thirty degrees, and then increases to the boiling-point.

166. Specific Heat.—Only one thirtieth as much heat is required to raise the temperature of a kilogram of mercury from zero to one

degree as is required to raise the temperature of a kilogram of water through the same range. In order to raise the temperatures of other substances through the same range, quantities of heat peculiar to each substance are required.

The quantity of heat required to raise the temperature of one kilogram of a substance from zero to one degree is called the *specific heat* of the substance.

If the temperature of one kilogram of a substance rise from t_0 to t, the limit of the ratio of the quantity of heat required to bring about the rise in temperature to the difference in temperature, as that difference diminishes indefinitely, is called the specific heat of the substance at temperature t. If we represent the quantity of heat by Q, the limit of the ratio $\frac{Q}{t - t_0} = \frac{dQ}{dt}$ expresses this specific heat.

The specific heats of substances are generally nearly constant between zero and one hundred degrees. The *mean specific heat* of a substance between zero and one hundred degrees is the one usually given in the tables.

The measurement of specific heat is one of the important objects of *calorimetry*.

167. Ice Calorimeter.—*Black's* or *Wilcke's ice calorimeter* consists of a block of pure ice having a cavity in its interior covered by a thick slab of ice. The body of which the specific heat is to be determined is heated to t degrees, then dropped into the cavity, and immediately covered by the slab. After a short time the temperature of the body falls to zero, and in so doing converts a certain quantity of ice into water. This water is removed by a sponge of known weight, and its weight is determined. It will be shown, that to melt a kilogram of ice requires 80 calories; if, then, the weight of the body be P, and its specific heat c, it gives up, in falling from t degrees to zero, Pct calories. On the other hand, if p kilograms of ice be melted, the heat required is $80p$. Therefore $Pct = 80p$; whence

$$c = \frac{80p}{Pt}. \qquad (61)$$

Bunsen's ice calorimeter (Fig. 66) is used for determining the specific heats of substances of which only a small quantity is at hand. The apparatus is entirely of glass. The tube *B* is filled with water and mercury, the latter extending into the graduated capillary tube *C*. To use the apparatus, alcohol which has been artificially cooled to a temperature below zero is passed through the tube *A*. A layer of ice forms around the outside of this tube. As water freezes, it expands. This causes the mercury to advance in the capillary tube *C*. When a sufficient quantity of ice has been formed, the alcohol is removed from *A*, the apparatus is surrounded by melting snow or ice, and a small quantity of water is introduced, which soon falls in temperature to zero. The position of the mercury in *C* is now noted; and the substance the specific heat of which is to be determined, at the temperature of the surrounding air, is dropped into the water in *A*. Its temperature quickly falls to zero, and the heat which it loses is entirely employed in melting the ice which surrounds the tube *A*. As the ice melts, the mercury in the tube *C* retreats. The change of position is an indication of the quantity of ice melted, and the quantity of ice melted measures the heat given up by the substance. The number of divisions of the tube *C* corresponding to one calorie can be determined by direct experiment.

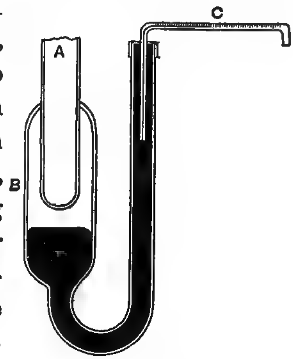

FIG. 66.

168. Method of Mixtures.—The method of mixtures consists in bringing together, at different temperatures, the substance of which the specific heat is desired and another of which the specific heat is known, and noting the change of temperature which each undergoes.

The *water calorimeter* consists of a vessel of very thin copper or brass, highly polished, and placed within another vessel upon non-conducting supports. A mass *P* of the substance of which the specific heat is to be determined is brought to a temperature t' in a suitable bath, then plunged in water at a temperature t, con-

tained in the calorimeter. The whole will soon come to a common temperature θ. The heat lost by the substance is $Pc(t'-\theta)$ calories. The heat gained by the calorimeter is the sum of that gained by the water and that gained by the materials of which the calorimeter is constructed. If p represent the mass of water, and p' the *water equivalent* of the calorimeter, or the mass of water which will rise by the same temperature interval as the calorimeter vessel does on the introduction of a given quantity of heat, the total heat gained by the calorimeter is $(p+p')(\theta-t)$; and hence

$$Pc(t'-\theta) = (p+p')(\theta-t), \tag{62}$$

from which c may be determined. The water equivalent p' is determined by experiment.

There is a source of error in the use of the instrument, due to the radiation of heat during the experiment. This error may be nearly eliminated by making a preliminary experiment to determine what change of temperature the calorimeter will experience; then, for the final experiment, the calorimeter and its contents are brought to a temperature below the temperature of the surrounding air, by about half the amount of that change. The calorimeter will then receive heat from the surrounding medium during the first part of the experiment, and lose heat during the second part. The rise of temperature is, however, much more rapid at the beginning than at the end of the experiment. The rise from the initial temperature to the temperature of the surrounding medium occupies less time than the rise from the latter to the final temperature. The gain of heat, therefore, does not exactly compensate for the loss. If greater accuracy be required, the rate of cooling of the calorimeter must be determined by putting into it warm water, the same in quantity as would be used in experiments for determining specific heat, and noting its temperature from minute to minute. Such an experiment furnishes the data for computing the loss or gain by radiation. To secure accurate results the body must be transferred from the bath to the calorimeter without sensible loss of heat.

169. Method of Comparison.—The method of comparison consists in conveying to the substance of which the specific heat is to be determined a known quantity of heat, and comparing the consequent rise of temperature with that produced by the same amount of heat in a substance of which the specific heat is known. In the early attempts to use this method, the heat produced by the same flame burning for a given time was applied successively to different liquids. A more exact method was the combustion, within the calorimeter, of a known weight of hydrogen. The best method of obtaining a known quantity of heat is by means of an electrical current of known strength flowing through a wire of known resistance wrapped upon the calorimeter.

170. Method of Cooling.—The method of cooling consists in noting the time required for the calorimeter, in a space kept constantly at zero, to cool from a temperature t' to a temperature t, when empty, when containing a given weight of water, and when containing a given weight of the substance of which the specific heat is sought. The *thermo-calorimeter* of Regnault, represented in Fig. 67, is an example. It consists of an alcohol thermometer, with its bulb A enlarged and made in the form of a hollow cylinder, inside of which the substance is placed. The thermometer is warmed, and then placed in a vessel surrounded by melting ice. It radiates heat to the sides of the vessel, and the column of alcohol in the tube falls. Let x be the time occupied in falling from the division n to the division n' when the space B is empty. Let the times occupied in falling between the same two divisions, when the space B contains a mass P of water, and when it contains a mass P' of the substance of which the specific heat c' is sought, be respectively x' and x''.

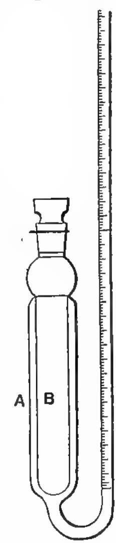

FIG. 67.

Let M be the water equivalent of the instrument. We then have $\frac{M}{x} = \frac{M + P}{x'} = \frac{M + P'c'}{x''}$, since, under the conditions of the experiment, the heat lost per second must be the same in each case.

Eliminating M, we obtain

$$c' = \frac{P}{P'}\left(\frac{x'' - x}{x' - x}\right). \tag{63}$$

171. Determination of the Mechanical Equivalent of Heat.—It has been stated that whenever heat is produced by the expenditure of mechanical energy, the quantity of heat produced is always proportional to the quantity of mechanical energy expended.

The *mechanical equivalent* of heat is the energy in mechanical units, the expenditure of which produces the unit of heat.

Heat applied to a body may increase the motion of its molecules; that is, add to their kinetic energy. It may perform internal work by moving the molecules against molecular forces. It may perform external work by producing motion against external forces. If we could estimate these effects in mechanical units, we might obtain the mechanical equivalent of heat. But the kinetic energy of the molecules cannot be estimated, for we do not know their mass or their velocity. We must, therefore, in the present state of our knowledge, resort to direct experiment to determine the heat equivalent. In one of the experiments of Joule, already referred to, a paddle-wheel was made to revolve, by means of weights, in a vessel filled with water. In this vessel were stationary wings, to prevent the water from acquiring a rotary motion with the paddle-wheel. By the revolution of the wheel the water was warmed. The heat so generated was estimated from the rise of temperature, while the mechanical energy required to produce it was given by the fall of the driving weight. Joule repeated this experiment, substituting mercury for the water. In another experiment he substituted an iron plate for the paddle-wheel, and made it revolve with friction upon a fixed iron plate under water.

Joule expressed his results in kilogram-metres—that is, the work done by a kilogram in falling under the force of gravity through one metre. He stated the mechanical equivalent of one calorie, in this unit, to be 423.9, from the experiments with water; 425.7, from those with mercury; and 426.1, from those with iron

plates. He gave the preference to the smallest value, and it has been generally accepted as the mechanical equivalent. This mechanical equivalent is called *Joule's equivalent,* and is represented by J. In absolute units, according to the later determinations of Griffiths, it is about 41,982,000,000 ergs per calorie.

Rowland has repeated Joule's experiment with water; but he caused the paddle-wheel to revolve by means of an engine, and determined the moment of the couple required to prevent the revolution of the calorimeter. Fig. 68 shows the apparatus. The shaft

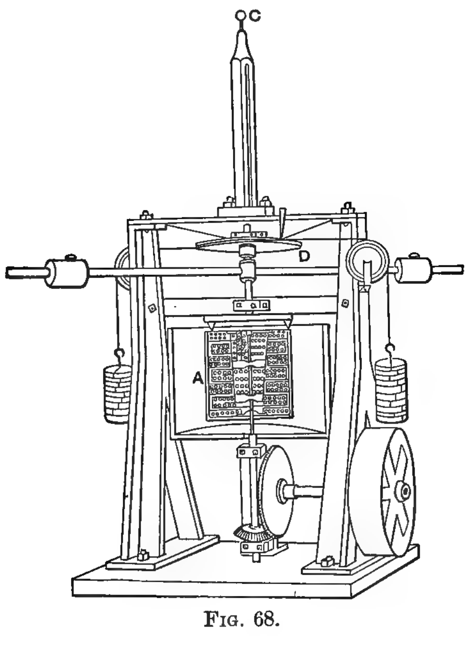

FIG. 68.

of the paddle-wheel projects through the bottom of the calorimeter, and is driven by means of a bevel-gear. The vessel A is suspended from C by a torsion wire, and its tendency to rotate balanced by weights attached to cords which act upon the circumference of a pulley D. By this disposition of the apparatus he was able to expend about one half a horse-power in the calorimeter, and obtain a

rise of temperature of 35° per hour; while in Joule's experiments the rise of temperature per hour was less than 1°. These experiments give, for the mechanical equivalent of one calorie at 5°, 429.8 kilogram-metres; at 20°, 426.4 kilogram-metres.

Several other methods have been employed for determining the mechanical equivalent. The concordance of the results by all these methods is sufficient to warrant the statement that the expenditure of a given amount of mechanical energy always produces the same amount of heat.

An experiment to determine the mechanical work done by the expenditure of a known quantity of heat was executed by Hirn. By the help of Regnault's measurements of the heat of vaporization Hirn was able to calculate the amount of heat which entered the cylinder, during the operation of a steam-engine, with the steam from the boiler, and by direct measurements he determined the amount of heat which left the cylinder during the operation of the engine and entered the condenser. So long as the engine was running without doing any external work, he found that these amounts of heat were appreciably equal; when the engine was made to do work, less heat passed from the cylinder into the condenser than had entered it from the boiler. A comparison of the amount of heat lost with the work done by the engine showed the same ratio between heat and work as that determined by Joule. Hirn's experiments were on so large a scale and the sources of error and the difficulties connected with the experiments were so numerous, that the number obtained by him for the mechanical equivalent of heat is of no great value. His experiments are, however, of very great interest because, while the experiments of Joule and of all the others who have worked on the problem prove the convertibility of work into heat, those of Hirn alone have proved the converse convertibility of heat into work.

CHAPTER II.

TRANSFER OF HEAT.

172. Transfer of Heat.—In the preceding discussions it has been assumed that heat may be transferred from one body to another, and that if two bodies in contact be at different temperatures, heat will be transferred from the hotter to the colder body. In general, if transfer of heat be possible in any system, heat will pass from the hotter to the colder parts of the system, and the temperature of the system will tend to become uniform. There are three ways in which this transfer is accomplished, called respectively convection, conduction, and radiation.

173. Convection.—If a vessel containing any fluid be heated at the bottom, the bottom layers become less dense than those above, producing a condition of instability. The lighter portions of the fluid rise, and the heavier portions from above, coming to the bottom, are in their turn heated. Hence continuous currents are caused. This process is called *convection.* By this process, masses of fluid, although fluids are poor conductors, may be rapidly heated. Water is often heated in a reservoir at a distance from the source of heat by the circulation produced in pipes leading to the source of heat and back. The winds and the great currents of the ocean are convection currents. An interesting result follows from the fact that water has a maximum density (§ 190). When the water of lakes cools in winter, currents are set up and maintained, so long as the surface water becomes more dense by cooling, or until the whole mass reaches 4°. Any further cooling makes the surface water lighter. It therefore remains at the surface, and its temperature

rapidly falls to the freezing-point, while the great mass of the water remains at the temperature of its maximum density.

174. Conduction.—If one end of a metal rod be heated, it is found that the heat travels along the rod, since those portions at a distance from the source of heat finally become warm. This process of transfer of heat from molecule to molecule of a body, while the molecules themselves retain their relative places, is called *conduction.*

In the discussion of the transfer of heat by conduction it is assumed as a principle, borne out by experiment, that the flow of heat between two very near parallel planes, drawn in a substance, is proportional to the difference of temperature between those planes, or that the flow of heat across a plane is proportional to the rate of fall of temperature across that plane.

175. Flow of Heat across a Wall.—The simplest body in which the flow of heat can be studied is a wall of homogeneous material bounded by two parallel infinite planes, one of which is kept at the temperature t' and the other at the temperature t; we represent the distance between the planes or the thickness of the wall by d. We suppose that the flow of heat across this wall has continued so long that it has become steady, or that the temperatures at all points have assumed final values. Manifestly the temperature at all points in any plane parallel with the faces of the wall is the same, and the same amount of heat passes across any one such plane as passes across any other. We conclude therefore by the fundamental principle assumed (§ 174) that the rate of change of temperature across each plane in the wall is the same, or that the change of temperature throughout the wall from one face to the other is uniform; the rate of change of temperature is therefore given by $\frac{t'-t}{d}$, where it has been assumed that t' is the higher temperature. If d' represent the distance of any plane in the wall from the hotter surface, the fall of temperature between it and the hotter surface is $(t'-t)\frac{d'}{d}$,

and the temperature of the plane is $t' - (t' - t)\frac{d'}{d}$. The confirmation by experiment of this law of temperature distribution in a wall is a warrant for our assumption of the fundamental principle of the flow of heat.

176. Conductivity.—If, now, we consider a prism extending across the wall, bounded by planes perpendicular to the exposed surfaces, and represent the area of its exposed bases by A, the quantity of heat which flows in a time T through this prism may be represented by

$$Q = K\frac{t' - t}{d}AT, \tag{64}$$

where K is a constant depending upon the material of which the wall is composed. K is the *conductivity* of the substance, and may be defined as the quantity of heat which in unit time flows through a section of unit area in a wall of the substance whose thickness is unity, when its exposed surfaces are maintained at a difference of temperature of one degree; or, in other words, it is the quantity of heat which in unit time flows through a section of unit area in a substance, where the rate of fall of temperature at that section is unity. In the above discussions the temperatures t' and t are taken as the actual temperatures of the surfaces of the wall. If the colder surface of the wall be exposed to air of temperature T, to which the heat which traverses it is given up, t will be greater than T. The difference will depend upon the quantity of heat which flows, and upon the facility with which the surface parts with heat.

177. Flow of Heat along a Bar.—If a prism of a substance have one of its bases maintained at a temperature t, while the other base and the sides are exposed to air at a lower temperature, the conditions of uniform fall of temperature no longer exist, and the amount of heat which flows through the different sections is no longer the same; but the amount of heat which flows through any section is still proportional to the rate of fall of temperature at that

section, and is equal to the heat which escapes from the portion of the bar beyond the section.

178. Measurement of Conductivity.—A bar heated at one end furnishes a convenient means of measuring conductivity. In Fig. 69 let AB represent a bar heated at A. Let the ordinates aa', bb', cc', represent the excess of temperatures above the temperature of the air at the points from which they are drawn. These temperatures may be determined by means of thermometers inserted in cavities in the bar, or by means of a thermopile. Draw the curve $a'b'c'd'$ . . . through the summits of the ordinates. The inclination of this curve at any point represents the rate of fall of temperature at that point. The ordinates to the line $b'm$, drawn tangent to the curve at the point b', show what would be the temperatures at various points of the bar if the fall were uniform and at the same rate as at b'. It shows that, at the rate of fall at b', the bar would at m be at the temperature of the air; or, in the length bm, the fall of temperature would equal the amount represented by bb'. The rate of fall is, therefore, $\frac{bb'}{bm}$. If Q represent the quantity of heat passing the section at b in the unit time, we have, from § 176,

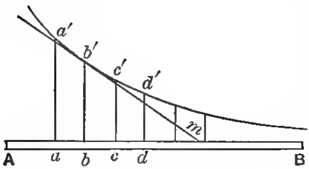

FIG. 69.

$$Q = K \times \textit{rate of fall of temperature} \times \textit{area of section}.$$

Q is equal to the quantity of heat that escapes in unit time from all that portion of the bar beyond b. It may be found by heating a short piece of the same bar to a high temperature, allowing it to cool under the same conditions that surround the bar AB, and observing its temperature from minute to minute as it falls. These observations furnish the data for computing the quantity of heat which escapes per minute from unit length of the bar at different temperatures. It is then easy to compute the amount of heat that escapes per minute from each portion, bc, cd, etc., of the bar beyond b; each portion being taken so short that its tempera-

ture throughout may, without sensible error, be considered uniform and the same as that at its middle point. Summing up all these quantities, we obtain the quantity Q which passes the section b in the unit time. Then

$$K = \frac{Q}{\textit{rate of fall of temperature at } b \times \textit{area of section}}.$$

179. Conductivity diminishes as Temperature rises.—By the method described above, Forbes determined the conductivity of a bar of iron at points at different distances from the heated end, and found that the conductivity is not the same at all temperatures, but is greater as the temperature is lower.

180. Conductivity of Crystals.—The conductivity of crystals of the isometric system is the same in all directions, but in crystals of the other systems it is not so. In a crystal of Iceland spar the conductivity is greatest in the direction of the axis of symmetry, and equal in all directions in a plane at right angles to that axis.

181. Conductivity of Non-homogeneous Solids.—De la Rive and De Candolle were the first to show that wood conducts heat better in the direction of the fibres than at right angles to them. Tyndall, by experimenting upon cubes cut from wood, has shown that the conductivity has a maximum value parallel to the fibres, a minimum value at right angles to the fibres and parallel to the annual layers. Feathers, fur, and the materials of clothing are poor conductors because of their want of continuity.

182. Conductivity of Liquids.—The conductivity of liquids can be measured, in the same way as that of solids, by noting the fall of temperature at various distances from the source of heat in a column of liquid heated at the top. Great care must be taken in these experiments to avoid errors due to convection currents.

Liquids are generally poor conductors.

183. Radiation.—We have now considered those cases in which there is a transfer of heat between bodies in contact. Heat is also transferred between bodies not in contact. This is effected by a process called *radiation,* which will be subsequently considered.

CHAPTER III.

EFFECTS OF HEAT.

184. The Kinetic Theory of Heat.—In order to describe more easily certain of the effects of heat, it is advantageous to have an idea of the theory by which they are explained. This theory, the *kinetic theory* of heat, asserts that the molecules of all bodies are in constant motion, and that the heat of a body is the kinetic energy of its molecules. The idea that heat consists of the motion of the least parts of matter was introduced into science by Newton, of course with a very imperfect knowledge of the facts. The apparently unlimited production of heat by mechanical work led Rumford and Davy, more particularly the latter, to assert the equivalence of heat and motion. This theory was afterwards displaced for many years by the influence of the French school of physicists, who considered bodies, at least in their mathematical discussions, as assemblages of stationary particles, and heat as a separate substance. It was revived by Mohr, who showed its very general applicability in the explanation of ordinary heat phenomena. Since the discovery of the conservation of energy, the reasons in its favor have been very much strengthened and its foundations securely laid by the complete success attained with it in explaining the laws of gases.

We will use this theory in its general form in the description of some of the effects of heat, and will discuss it more fully in § 221 *seq.*

SOLIDS AND LIQUIDS.

185. Expansion of Solids.—When heat is applied to a body it increases the kinetic energy of the molecules, and also increases the

potential energy, by forcing the molecules farther apart against their mutual attractions and any external forces that may resist expansion. Since the internal work to be done when a solid or liquid expands varies greatly for different substances, it might be expected that the amount of expansion for a given rise of temperature would vary greatly.

In studying the expansion of solids, we distinguish *linear* and *voluminal* expansion.

The increase which occurs in the unit length of a substance for a rise of temperature from zero to 1° C. is called the *coefficient of linear expansion.* Experiment shows that the expansion for a rise of temperature of one degree is very nearly constant between zero and 100°.

Represent by l_0 the distance between two points in a body at zero, by l_t the distance between the same points at the temperature t, and by α the coefficient of linear expansion of the substance of which the body is composed.

The increase in the distance l_0 for a rise of one degree in temperature is αl_0, for a rise of t degrees $\alpha t l_0$. Hence we have, after a rise in temperature of t degrees,

$$l_t = l_0(1 + \alpha t). \tag{65}$$

The binomial $1 + \alpha t$ is called the *factor of expansion.*

In the same way, if k represent the *coefficient of voluminal expansion,* the volume of a body at a temperature t will be

$$V_t = V_0(1 + kt); \tag{66}$$

and if d represent density, since density is inversely as volume, we have

$$d_t = \frac{d_0}{1 + kt}. \tag{67}$$

For a homogeneous isotropic solid, the coefficient of voluminal expansion is three times that of linear expansion; for, if the temperature of a cube, with an edge of unit length, be raised one degree, the length of its edge becomes $1 + \alpha$, and its volume $1 + 3\alpha + 3\alpha^2 + \alpha^3$. Since α is very small, its square and cube

may be neglected; and the volume of the cube after a rise in temperature of one degree is $1 + 3\alpha$. 3α is, therefore, the coefficient of voluminal expansion.

186. Measurement of Coefficients of Linear Expansion.—Coefficients of linear expansion are measured by comparing the lengths, at different temperatures, of a bar of the substance the coefficient of which is required, with the length, at constant temperature, of another bar. The constant temperature of the latter bar is secured by immersing it in melting ice. The bar the coefficient of which is sought may be brought to different temperatures by immersing it in a liquid bath; but it is found better to place the bar upon the instrument by means of which the comparisons are to be made, and leave it for several hours exposed to the air of the room, which is kept at a constant temperature by artificial means. Of course several hours must elapse between any two comparisons by this method, and its application is restricted to such ranges of temperature as may be obtained in occupied rooms; but within this range the observations can be made much more accurately than would be the case when the bar is immersed in a bath, and it is within this range that an accurate knowledge of coefficients of expansion is of most importance.

187. Expansion of Liquids.—In studying the expansion of a liquid, it is important to distinguish its *absolute* expansion, or the real increase in volume, and its *apparent* expansion, or its increase in volume in comparison with that of the containing vessel.

To determine the absolute expansion, some method must be used which does not require a knowledge of the expansion of the vessel containing the liquid. The method used by Regnault in determining the absolute expansion of mercury consisted in comparing the heights of two columns of mercury at different temperatures when they were so adjusted as to give the same pressure.

The apparent expansion is determined by filling a vessel of known volume with the liquid at one temperature, and by measuring the amount of the liquid which runs out when the temperature is raised. This method was also used by Regnault in his study of

the expansion of mercury. The vessel which he used was a glass bulb furnished with a capillary tube. It was filled with mercury at a known temperature, and its volume determined by the weight of the mercury contained in it and the specific gravity of mercury. It was then heated to another known temperature, and the mercury which ran out was collected and weighed. From these data the apparent expansion of mercury in glass could be determined.

When the absolute expansion of mercury is known the knowledge of its apparent expansion in glass enables us to determine the absolute expansion of glass also.

If the apparent expansion of mercury be known, and if we assume that its expansion is proportional to the rise of temperature, we may evidently use the amount of mercury which runs out when the bulb is heated as a measure of its change of temperature. The instrument just described is therefore called a *weight thermometer*.

188. Determination of Voluminal Expansion of Solids.—The weight thermometer may be used to determine the coefficient of voluminal expansion of solids. For this purpose, the solid, of which the volume at zero is known, must be introduced into the bulb by the glass-blower. If the bulb containing the solid be filled with mercury at zero, and afterward heated to the temperature t, it is evident that the amount of mercury that will overflow will depend upon the coefficient of expansion of the solid, and upon the coefficient of apparent expansion of mercury. If the latter has been determined for the kind of glass used, the former can be deduced. By this means the coefficients of voluminal expansion of some solids have been determined; and the results are found to verify the conclusion, deduced from theory (§ 185), that the voluminal coefficient is three times the linear.

189. Absolute Expansion of Liquids other than Mercury.—The weight-thermometer may also serve to determine the coefficients of expansion of liquids other than mercury; for, if the absolute expansion of glass has been found as described above, the instrument may be filled with the liquid the coefficient of which is desired, and the

apparent expansion of this liquid found exactly as was that of mercury. The absolute coefficient for the liquid is then the sum of the coefficient of apparent expansion and the coefficient for the glass.

190. Expansion of Water.—The use of water as a standard with which to compare the densities of other substances makes it necessary to know, not merely its mean coefficient of expansion, but its actual expansion, degree by degree. This is the more important since water expands very irregularly. The best determinations of the volumes of water at different temperatures are those of Matthiessen. The method which he employed was to weigh in water a mass of glass of which the coefficient of expansion had been previously determined.

Water contracts, instead of expanding, from 0° to 4°. At 4° it is at its maximum density, and from that temperature to its boiling-point it expands.

191. Effect of Variation of Temperature upon Specific Heat.—It has already been stated (§ 166) that the specific heat of bodies changes with temperature. With most substances the specific heat increases as the temperature rises.

For example, the true specific heat of the diamond

At 0° is	0.0947
At 50° is	0.1435
At 100° is	0.1905
At 200° is	0.2719

192. Effect of Change of Physical State upon Specific Heat.—The specific heat of a substance is not the same in its different physical states. In the solid or gaseous state of the substance it is generally less than in the liquid. For example:

	Mean Specific Heat.		
	Solid.	Liquid.	Gaseous.
Water	0.504	1.000	0.481
Mercury	0.0314	0.333	
Tin	0.056	0.0637	
Lead	0.0314	0.0402	
Bromine		0.1129	0.0555

193. Dulong and Petit's Law. Atomic Heat.—In their study of the specific heats of a number of chemical elements which are solid at ordinary temperatures, Dulong and Petit found that the product of the specific heat by the atomic weight of the element was approximately a constant quantity. Further researches, especially those of Kopp, have confirmed this statement as a general law for all solid elements. The constant number to which the product of the specific heat and the atomic weight approximates is ordinarily given as 6.4 when the specific heat is measured in calories, though this is probably a little too high. The deviations from this number presented by different elements are rather large, amounting in many cases to as much as 5 per cent.

If masses of the different elements be taken which are proportional to their atomic weights, these masses will contain the same numbers of atoms. The heat required to raise one of these masses one degree in temperature is therefore the same for all such substances. This statement is of course true only within the limits of accuracy with which the different substances conform to Dulong and Petit's law. The experiments of F. Neumann and Regnault showed that a similar law applies to compounds of solids which are of the same chemical constitution; that is, which contain the same number of atoms in the molecule. For all such bodies the product of the specific heat and the molecular weight is a constant; this constant is different for the different classes of substances—that is, for those substances which have different numbers of atoms in the molecule. But if the constant obtained for each class of substances be divided by the number of atoms in the molecule of that class, the quotient is approximately the same constant, 6.4, as that obtained for the elements. By applying this law to compounds in which one of the elements is a substance, like hydrogen, which cannot be examined directly in the solid state, the atomic heat of that substance may be calculated. It is found that the atomic heats of certain substances, notably hydrogen, carbon, oxygen, nitrogen, and silicium, deviate very widely from the constant with which the other atomic heats approximately agree.

The elementary gases obey a similar law with considerable exactness; the constant given by the product of their specific heats at constant pressure and their atomic weights is about 3.4.

The following table will illustrate the law of Dulong and Petit. The atomic weights are those given by Clarke.

Elements.	Specific Heat of Equal Weights.	Atomic Weight.	Product of Specific Heat and Atomic Weight.
Iron........	0.114	55.9	6.372
Copper	0.095	63.17	6.001
Mercury....	0.0314 (solid)	199.71	6.128
Silver.......	0.057	107.67	6.137
Gold.......	0.0329	196.15	6.453
Tin	0.056	117.7	6.591
Lead.......	0.0314	206.47	6.483
Zinc........	0.0955	64.9	6.198

194. Fusion and Solidification.—When ice at a temperature below zero is heated, its temperature rises to zero, and then the ice begins to melt; and, however high the temperature of the medium that surrounds it may be, its temperature remains constant at zero so long as it remains in the solid state. This temperature is the *melting-point of ice*, and because of its fixity it is used as one of the standard temperatures in graduating thermometric scales. Other bodies melt at very different but at fixed and definite temperatures. Many substances cannot be melted, as they decompose by heat.

Alloys often melt at a lower temperature than any of their constituents. An alloy of one part lead, one part tin, four parts bismuth, melts at 94°; while the lowest melting-point of its constituents is that of tin, 228°. An alloy of lead, tin, bismuth, and cadmium melts at 62°.

If a liquid be placed in a medium the temperature of which is below its melting-point, it will in general begin to solidify when its temperature reaches its melting-point, and it will remain at that temperature until it is all solidified. Under certain conditions, however, the temperature of a liquid may be lowered several degrees below its melting-point without solidification, as will be seen below.

195. Change of Volume with Change of State.—Substances are generally more dense in the solid than in the liquid state, but there are some notable exceptions. Water, on solidifying, expands; so that the density of ice at zero is only 0.9167, while that of water at 4° is 1. This expansion exerts considerable force, as is evidenced by the bursting of vessels and pipes containing water.

196. Change of Melting- and Freezing-points.—If water be enclosed in a vessel sufficiently strong to prevent its expansion, it cannot freeze except at a lower temperature. The freezing-point of water is, therefore, lowered by pressure. On the other hand, substances which contract on solidifying have their solidification hastened by pressure.

The lowering of the melting-point of ice by pressure explains some remarkable phenomena. If pieces of ice be pressed together, even in warm water, they will be firmly united. Fragments of ice may be moulded under heavy pressure, into a solid, transparent mass. This soldering together of masses of ice is called *regelation.* If a loop of wire be placed over a block of ice and weighted, it will cut its way slowly through the ice, and regelation will occur behind it. After the wire has passed through, the block will be found one solid mass, as before. The explanation of these phenomena is, that the ice is partially melted by the pressure. The liquid thus formed is colder than the ice; it finds its way to points of less pressure, and there, because of its low temperature, it congeals, firmly uniting the two masses.

Water, when freed from air and kept perfectly quiet, will not form ice at the ordinary freezing-point. Its temperature may be lowered to −10° or −12° without solidification. In this condition a slight jar, or the introduction of a small fragment of ice, will cause a sudden congelation of part of the liquid, accompanied by a rise in temperature in the whole mass to zero.

A similar phenomenon is observed in the case of several solutions, notably sodium sulphate and sodium acetate. If a saturated hot solution of one of these salts be made, and allowed to cool in a closed bottle in perfect quiet, it will not crystallize. Upon opening

the bottle and admitting air, crystallization commences, and spreads rapidly through the mass, accompanied by a considerable rise of temperature. If the amount of salt dissolved in the water be not too great, the solution will remain liquid when cooled in the open air, and it may even suffer considerable disturbance by foreign bodies without crystallization; but crystallization begins immediately upon contact with the smallest crystal of the same salt.

197. Freezing-point of Solutions.—It has been long known that the freezing-point of a solution of salt and water is lower than that of pure water. The relation of the lowering of the freezing-point to the concentration of the solution was investigated by Blagden, who found that for dilute solutions the lowering of the freezing-point was proportional to the concentration. This matter has been investigated by Raoult, who established some most important generalizations. Raoult showed that, for indifferent solutions, that is, for solutions which are not electrolytes, provided they are very dilute, the lowering of the freezing-point is very closely proportional to the concentration; its amount differs for different solvents. He further showed that, for any one solvent, the lowering of the freezing-point is the same whatever be the dissolved substance, provided that the solutions are *equimolecular*, that is, contain the same number of molecules of the dissolved substance in unit volume of the solution. It may be shown on theoretical grounds that the change in the freezing-point depends upon the osmotic pressure, the freezing-point, the heat of fusion, and the density of the solution.

Solutions which are electrolytes, or are not indifferent, also exhibit a lowering of the freezing-point proportional to the concentration, but the amount of change is greater than in indifferent solutions. This difference is explained by assuming a partial or complete dissociation of the molecules of the dissolved substances into their constituent ions (§ 285).

198. Heat Equivalent of Fusion.—Some facts that have appeared in the above account of the phenomena of fusion and solidification require further study. It has been seen that, however rapidly the temperature of a solid may be rising, the moment fusion

begins the rise of temperature ceases. Whatever the heat to which a solid may be exposed, it cannot be made hotter than its melting-point. When ice is melted by pressure its temperature is lowered. When a liquid is cooled, its fall of temperature ceases when solidification begins; and if, as may occur under favorable conditions, a liquid is cooled below its melting-point, its temperature rises at once to the melting-point, when solidification begins. Heat, therefore, disappears when a body melts, and is generated when a liquid becomes solid.

It was stated (§ 159) that ice can be melted by friction; that is, by the expenditure of mechanical energy. Fusion is, therefore, work which requires the expenditure of some form of energy to accomplish it. The heat required to melt unit mass of a substance is the *heat equivalent of fusion* of that substance. When a substance solidifies, it develops the same amount of heat as was required to melt it.

As will be shown later at greater length, the absorption of heat which occurs when a solid is melted is explained by supposing that it is used in doing work against the forces which determine the direction of the molecules in the solid and in increasing the kinetic energy of molecular translation.

199. Determination of the Heat Equivalent of Fusion.—The heat equivalent of fusion may be determined by the method of mixtures (§ 168), as follows: A mass of ice, for example, represented by P, at a temperature t below its melting-point, to insure dryness, is plunged into a mass P' of warm water at the temperature T. Represent by θ the resulting temperature, when the ice is all melted. If p represent the water equivalent of the calorimeter, $(P' + p)(T - \theta)$ is the heat given up by the calorimeter and its contents. Let c represent the specific heat of ice, and x the heat equivalent of fusion. The ice absorbs, to raise its temperature to zero, Ptc calories; to melt it, Px calories; to warm the water after melting, $P\theta$ calories. We then have the equation

$$Ptc + P\theta + Px = (P' + p)(T - \theta), \tag{68}$$

from which x may be found.

Other calorimetric methods may be employed. The best experiments give, for the heat equivalent of fusion of ice, very nearly eighty calories.

VAPORS AND GASES.

200. The Gaseous State.—A *gas* may be defined as a highly compressible fluid. A given mass of gas has no definite volume. Its volume varies with every change in the external pressure to which it is exposed. A *vapor* is the gaseous state of a substance which at ordinary temperatures exists as a solid or a liquid.

201. Vaporization is the process of formation of vapor. There are two phases of the process: *evaporation*, in which vapor is formed at the free surface of the liquid; and *ebullition*, in which the vapor is formed in bubbles in the mass of the liquid, or at the heated surface with which it is in contact.

202. Evaporation.—If a liquid be enclosed in a vessel which it does not entirely fill, the space above the liquid begins at once to be occupied by the vapor of the liquid. The presence of the vapor can be detected in many ways, some of which are applicable only in special cases. Those which are always applicable are the measurement of the increased pressure due to the vapor and the condensation of the vapor into the liquid state after isolating it from the mass of liquid beneath it. The process of forming vapor in this way is *evaporation*. Evaporation goes on continually from the free surfaces of many liquids, and even of solids. It increases in rapidity as the temperature increases, and ceases when the vapor has reached a certain density, always the same for the same temperature, but greater for a higher temperature. It goes on very rapidly in a vacuum; but it is found that the final density of the vapor is no greater, or but little greater, than when some other gas is present. While evaporation is going on, heat must be supplied to the liquid to keep its temperature constant.

Evaporation may be readily explained on the kinetic theory (§ 184) on the supposition that, in the interaction of the molecules, the motion of any one may be more or less violent, as it receives

motion from its neighbors or gives up motion to them. At the exposed surface of the substance the motion of a molecule may at times be so violent as to project it beyond the reach of the molecular attractions. If this occur in the air, or in a space filled with any gas, the molecule may be turned back, and made to rejoin the molecules in the liquid mass; but many will find their way to such a distance that they will not return. They then constitute a vapor of the substance. As the number of free molecules in the space above the liquid increases, it is plain that there may come a time when as many will rejoin the liquid as escape from it. The space is then *saturated* with the vapor. The more violent the motion in the liquid, that is, the higher its temperature, the more rapidly the molecules will escape, and the greater must be the number in the space above the liquid before the returning will equal in number the outgoing molecules. In other words, the higher the temperature, the more dense the vapor that saturates a given space. If the space above a liquid be a vacuum, the escaping molecules will at first meet with no obstruction, and, as a consequence, the space will be very quickly saturated with the vapor. The presence of another vapor or a gas impedes the motion of the outgoing molecules, and causes evaporation to go on slowly, but it has very little influence upon the number of molecules that must be present in order that those which return may equal in number those which escape. Since only the more rapidly moving molecules escape, they carry off more than their share of the heat of the liquid, and thus the temperature will fall unless heat is supplied from without.

203. Pressure of Vapors.—As a liquid evaporates in a closed space, the vapor formed exerts a pressure upon the enclosure and upon the surface of the liquid, which increases so long as the quantity of vapor increases, and reaches a maximum when the space is saturated. This *maximum pressure* of a vapor increases with the temperature. When evaporation takes place in a space filled by another gas which has no action upon the vapor, the pressure of the vapor is added to that of the gas, and the pressure of the mixture is, therefore, the sum of the pressures of its constituents. The law

was announced by Dalton that the quantity of vapor which saturates a given space, and consequently the maximum pressure of that vapor, is the same whether the space be empty or contain a gas. Regnault has shown that, for water, ether, and some other substances, the maximum pressure of their vapors is slightly less when air is present.

204. The Vapor Pressure of Solutions.—The pressure of the saturated vapor formed from an indifferent solution, or one which is not an electrolyte, is always less than the vapor pressure of the pure solvent. Raoult discovered that the diminution of vapor pressure is proportional to the concentration, provided the solutions are very dilute and that, for any one solvent, the diminution of vapor pressure is the same, whatever be the dissolved substance, provided the solutions are equimolecular, that is, contain the same number of molecules in equal volumes of the solutions. It may be shown on theoretical grounds that the diminution of vapor pressure depends upon the density of the vapor and the osmotic pressure and density of the solution.

Solutions which are not electrolytes, or which are not indifferent, exhibit a diminution of vapor pressure proportional to the concentration, but the amount of change is greater than in indifferent solutions. This difference is explained by assuming a partial or complete dissociation of the molecules of the dissolved substances into their constituent ions (§ 285).

205. Ebullition.—As the temperature of a liquid rises, the pressure which its vapor may exert increases, until a point is reached where the vapor is capable of forming, in the mass of the liquid, bubbles which can withstand the superincumbent pressure of the liquid and the atmosphere above it. These bubbles of vapor, escaping from the liquid, give rise to the phenomenon called *ebullition*, or *boiling*. Boiling may, therefore, be defined as the agitation of a liquid by its own vapor.

Generally speaking, for a given liquid, ebullition always occurs at the same temperature for the same pressure; and, when once commenced, the temperature of the liquid no longer rises, no

matter how intense the source of heat. This fixed temperature is called the *boiling-point* of the liquid. It differs for different liquids, and for the same liquid under different pressures. That the boiling-point must depend upon the pressure is evident from the explanation of the phenomenon of ebullition above given.

Substances in solution, if less volatile than the liquid, raise the boiling-point. While pure water boils at 100°, water saturated with common salt boils at 109°. The material of the containing vessel also influences the boiling-point. In a glass vessel the temperature of boiling water is higher than in one of metal. If water be deprived of air by long boiling, and then cooled, its temperature may afterwards be raised considerably above the boiling-point before ebullition commences. Under these conditions the first bubbles of vapor will form with explosive violence. The air dissolved in water separates from it at a high temperature in minute bubbles. Into these the water evaporates, and, whenever the elastic force of the vapor is sufficient to overcome the superincumbent pressure, it enlarges them, and causes the commotion that marks the phenomenon of ebullition. If no such openings in the mass of the fluid exist, the cohesion of the fluid, or its adhesion to the vessel, as well as the pressure, must be overcome by the vapor. This explains the higher temperature at which ebullition commences when the liquid has been deprived of air.

206. Spheroidal State.—If a liquid be introduced into a highly heated capsule, or poured upon a very hot plate, it does not wet the heated surface, but forms a flattened spheroid, which presents no appearance of boiling, and evaporates only very slowly. Boutigny has carefully studied these phenomena, and made known the following facts: The temperature of the spheroid is below the boiling-point of the liquid. The spheroid does not touch the heated plate, but is separated from it by a non-conducting layer of vapor. This accounts for the slowness of the evaporation. To maintain the liquid in this condition the temperature of the capsule must be much above the boiling-point of the liquid; for water it must be at least 200° C. If the capsule be allowed to cool, the temperature

will soon fall below the limit necessary to maintain the spheroidal state, the liquid will moisten the capsule, and there will be a rapid ebullition with disengagement of vapor. If a liquid of very low boiling-point, as liquid nitrous oxide, which boils at $-88°$, be poured into a red-hot capsule, it will assume the spheroidal state; and, since its temperature cannot rise above its boiling-point, water, or even mercury, plunged into it, will be frozen.

207. Production of Vapor in a Limited Space.—When a liquid is heated in a limited space the vapor generated accumulates, increasing the pressure, and the temperature rises above the ordinary boiling-point. Cagniard-Latour experimented upon liquids in spaces but little larger than their own volumes. He found that, at a certain temperature, the liquid suddenly disappeared; that is, it was converted into vapor in a space but little larger than its own volume. It is supposed that above the temperature at which this occurs, which is called the *critical temperature*, the substance cannot exist in the liquid state (§ 223).

208. Liquefaction.—Only a certain amount of vapor can exist at a given temperature in a given space. If the temperature of a space saturated with vapor be lowered, some of the vapor must condense into the liquid state. It is not necessary that the temperature of the whole space be lowered; for when the vapor in the cooled portion is condensed, its pressure is diminished, the vapor from the warmer portion flows in, to be in its turn condensed, and this continues until the whole is brought to the density and pressure due to the cooled portion. Any diminution of the space occupied by a saturated vapor at constant temperature will cause some of the vapor to become liquid, for, if it do not condense, its pressure must increase; but a saturated vapor is already at its maximum pressure.

If the vapor in a given space be not at its maximum pressure, its pressure will increase when its volume is diminished, until the maximum pressure is reached; when, if the temperature remain constant, further reduction of volume causes condensation into the liquid state, without further increase of pressure or density. This statement is true of several of the gases at ordinary temperatures.

Chlorine, sulphur dioxide, ammonia, nitrous oxide, carbon dioxide, and several other gases, become liquid under sufficient pressure. Andrews found that at a temperature of 30.92° pressure ceases to liquefy carbon dioxide. This is the critical temperature for that substance. The critical temperatures of oxygen, hydrogen, and the other so-called permanent gases, are so low that it is only by methods capable of yielding an extremely low temperature that they can be liquefied. By the use of such methods any of the gases may be made to assume the liquid state. In the case of hydrogen, however, the low temperature necessary for its liquefaction has only been reached by allowing the gas to expand from a condition of great condensation, in which it had already been cooled to a very low point. The first successful attempts to condense these gases were made by Cailletet and Pictet, working independently. The best work on the subject has been done by Olszewski, who has succeeded in obtaining large quantities of liquid oxygen, nitrogen, and hydrogen, and in freezing nitrogen.

209. Pressure and Density of Saturated Gases and Vapors.—It has been seen that, for each gas or vapor at a temperature below the critical temperature, there is a maximum pressure which it can exert at that temperature. To each temperature there corresponds a maximum pressure, which is higher as the temperature is higher. A gas or vapor in contact with its liquid in a closed space will exert its maximum pressure.

The relation between the temperature and the corresponding maximum pressure of a vapor is a very important one, and has been the subject of many investigations. The vapor of water has been especially studied, the most extensive and accurate experiments being those of Regnault.

210. Pressure and Density of Non-saturated Gases and Vapors.—If a gas or vapor in the non-saturated condition be maintained at constant temperature, it follows very nearly Boyle's law (§ 105). If its temperature be below its critical temperature, the product of volume by pressure diminishes, and near the point of saturation the departure from the law may be considerable. At this point

the pressure becomes constant for any further diminution of volume, and the gas assumes the liquid state. The less the pressure and density of the gas, the more nearly it obeys Boyle's law.

211. Gay-Lussac's Law.—It has been stated already that gases expand as the temperature rises. The law of this expansion, called, after its discoverer, *Gay-Lussac's law*, is that, for each increment of temperature of one degree, every gas expands by the same constant fraction of its volume at zero. This is equivalent to saying that a gas has a constant coefficient of expansion, which is the same for all gases.

Let V_0, V_t represent the volumes at zero and t respectively, and α the coefficient of expansion. Then, the pressure remaining constant, we have

$$V_t = V_0(1 + \alpha t). \tag{69}$$

If d_0, d_t represent the densities at the same two temperatures we have, since densities are inversely as volumes,

$$d_t = \frac{d_0}{1 + \alpha t}. \tag{70}$$

Later investigations, especially those of Regnault, show that this simple law, like the law of Boyle, is not rigorously true, though it is very nearly so for all gases and vapors which are not too near their points of saturation. The common coefficient of expansion is $\alpha = 0.003666 = \frac{1}{273}$ very nearly.

212. Boyle's and Gay-Lussac's Laws.—From the law of Boyle we have, for a given mass of gas, if the temperature remain constant, $V_p p = V_{p'} p' =$ volume at pressure unity, where V_p, $V_{p'}$ represent the volumes at pressure p and p' respectively.

From the law of Gay-Lussac we have, if the pressure remain constant, $V_0 = \frac{V_t}{1 + \alpha t} = \frac{V_{t'}}{1 + \alpha t'}$. If the temperature and pressure both vary, we have

$$\frac{V_{pt} p}{1 + \alpha t} = \frac{V_{p't'} p'}{1 + \alpha t'}; \tag{71}$$

that is, if the volume of a given mass of gas be multiplied by the corresponding pressure and divided by the factor of expansion, the quotient is constant.

Let us represent this constant by C and write $\frac{1}{273}$ for α and v for V_{pt}. Then we have $\dfrac{vp}{273+t} = \dfrac{C}{273} = R$, where R is a constant. If the temperatures of the gas be reckoned from a zero point which is 273° below the melting-point of ice, or the zero of the centigrade thermometer, we may set $273 + t = T$, where T is the temperature reckoned from the new zero, and have finally

$$pv = RT \tag{72}$$

as the equation which embodies Boyle's and Gay-Lussac's laws. The temperature T is called the temperature on the scale of the air-thermometer, and the zero from which it is reckoned is called the zero of the air-thermometer. For reasons which will subsequently appear, it is also called the absolute temperature, and its zero the absolute zero.

213. Elasticity of Gases.—It has been shown (§ 105) that the elasticity of a gas obeying Boyle's law is numerically equal to the pressure. This is the *elasticity for constant temperature.* But when a gas is compressed it is heated (§ 158); and heating a gas increases its pressure. Under ordinary conditions, therefore, the ratio of a small increase of pressure to the corresponding decrease of unit volume is greater than when the temperature is constant. It is important to consider the case when all the heat generated by the compression is retained by the gas. The elasticity is then a maximum, and is called the *elasticity when no heat is allowed to enter or escape.*

Let mn (Fig. 70) be a curve representing the relation between volume and pressure for constant temperature, of which the abscissas represent volumes and the ordinates pressures. Such a curve is called an *isothermal* line. It is plain that to each temperature must correspond its own isothermal line. If, now, we suppose the gas to be compressed, and no heat to escape, it is plain

that if the volume diminish from OC to OG, the pressure will become greater than GD; suppose it to be GM. If a number of such points as M be found, and a line be drawn through them, it will represent the relation between volume and pressure when no heat enters or escapes. It is called an *adiabatic* line. It evidently makes a greater angle with the horizontal than the isothermal.

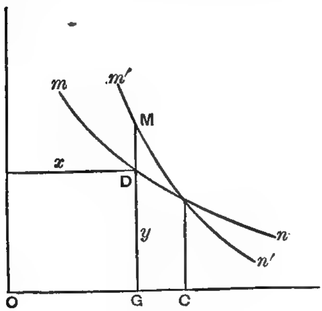

FIG. 70.

The tangents to these lines at the point of intersection, being the ratios of the changes of pressure to the same changes of volume under the conditions represented by those lines are proportional to the elasticity at constant temperature, or the isothermal elasticity E_t, and to the elasticity when no heat is allowed to enter or escape, or the adiabatic elasticity E_h, respectively.

214. Specific Heats of Gases.—The amount of heat necessary to raise the temperature of unit mass of a gas one degree, while the volume remains unchanged, is called *the specific heat of the gas at constant volume.* The amount of heat necessary to raise the temperature of unit mass of a gas one degree when expansion takes place without change of pressure, is called *the specific heat of the gas at constant pressure.*

The determination of the relation of these two quantities is a very important problem.

The specific heat of a gas at constant pressure may be found by passing a current of warmed gas through a tube coiled in a calorimeter. There are great difficulties in the way of an accurate determination, because of the small density of the gas, and the time required to pass enough of it through the calorimeter to obtain a reasonable rise of temperature. The various sources of error produce effects which are sometimes as great as, or even greater than, the quantity to be measured. It is beyond the scope of this work

to describe in detail the means by which the effects of the disturbing causes have been determined or eliminated.

The specific heat of a gas at constant volume is generally determined from the ratio between it and the specific heat at constant pressure. The first direct determination of this ratio was accomplished by Clement and Desormes. It is now most commonly determined from the velocity of sound (§§ 135, 216).

215. Work Done by the Expansion of a Gas.—It was shown by Joule that when a gas expands without doing external work, its temperature remains practically constant. His experiment consisted in allowing gas compressed within a reservoir to flow into another reservoir in which a vacuum had been made. The reservoirs were immersed in the water of a calorimeter; it was found that in these circumstances the expansion of the gas was not attended either by the evolution or absorption of heat. As the gas had done no external work during the expansion, this proved that its energy remained unchanged. The energy of a gas is therefore a function of its temperature alone.

If the temperature of a unit mass of gas be raised 1° while its volume is kept constant, the quantity of heat C_v, the specific heat at constant volume, must enter the gas. If its temperature be raised by the same amount while it is allowed to expand under constant pressure and to do work W by that expansion, a quantity of heat C_p, the specific heat at constant pressure, must be used. Since the gas is at the same temperature at the end of each of these operations, its energy must be the same in both cases, and the difference between the quantities of heat employed, or $C_p - C_v$, must be equal to the work W done by the expansion.

The experiments of Joule and Thomson, which proved that the experiment of Joule just described was not sufficiently sensitive to yield an exact result, and that the temperature of a gas really falls slightly when it expands without doing external work, do not seriously invalidate the conclusion just drawn; they merely prove that some internal work is done in the gas during its expansion. This internal work is so small in amount that it may be neglected in most cases.

216. Ratio of the Elasticities and of the Specific Heats of a Gas. —The ratio of the two principal specific heats of a gas is the same as the ratio of its two principal elasticities. To show this, construct an adiabatic line ϕ and an isothermal line θ (Fig. 71) intersecting at the point O; from that point draw a line parallel with the axis of volumes and take a point A on that line very near the point O. Through that point draw a line parallel with the axis of pressures, intersecting the isothermal and the adiabatic lines at B and C respectively.

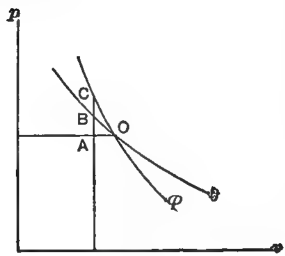

FIG. 71.

OA is the diminution of volume, Δv, caused by an increase of pressure $AB = \delta p$ if the compression is isothermal, or by the increase of pressure $AC = \Delta p$ if the compression is adiabatic. From the definition of elasticity (§ 102) we have the equations

$$E_t = \frac{v \,.\, \delta p}{\Delta v},\ E_h = \frac{v \,.\, \Delta p}{\Delta v},\ \text{and hence } \frac{E_h}{E_t} = \frac{\Delta p}{\delta p}.$$

We will now determine the value of the ratio $\frac{\Delta p}{\delta p}$ in terms of the principal specific heats. For convenience we assume that we are dealing with a unit mass of gas. The diminution of volume Δv at constant pressure sets free the quantity of heat $C_p \,.\, \Delta t$, where Δt is the change of temperature that occasions the change of volume; the point A then represents the condition of the gas. The gas may be brought into this same condition by an adiabatic compression from O to C, during which no heat either enters or leaves the gas, and by a diminution of pressure $AC = \Delta p$ while the volume is constant, caused by the abstraction of the heat produced by the compression. The heat which must be abstracted from the gas in order that it shall attain the condition denoted by A, is to the heat that must be abstracted to cause the diminution of pressure $BA = \delta p$ in the ratio of Δp to δp. The heat which must be abstracted to cause the diminution of pressure $BA = \delta p$ at constant volume is $C_v \,.\, \Delta t$, where Δt has the same value as before, since the

change of temperature is that experienced in passing from the isothermal OB to the isothermal which passes through A. The heat abstracted to produce the diminution of pressure Δp is therefore $C_v \cdot \frac{\Delta p}{\delta p} \cdot \Delta t$. Now the internal energy of the gas in the condition represented by A depends only on its temperature and is independent of the way in which that condition is reached. The work done on the gas in its change from O to A does depend on the way in which the change is effected, but the difference between the work done on it during the first operation and that done on it during the second operation is an infinitesimal of the second order, represented by the area OCA (§ 232), and may be neglected. The quantities of heat abstracted during the two operations may therefore be set equal, so that we have $C_p \Delta t = C_v \frac{\Delta p}{\delta p} \Delta t$, and hence

$$\frac{\Delta p}{\delta p} = \frac{C_p}{C_v} = \frac{E_h}{E_t}, \tag{73}$$

by the equation already obtained.

It has been shown that the velocity of sound in any medium is equal to the square root of the quotient of the elasticity divided by the density of the medium; that is, *velocity* $= \sqrt{\frac{E}{D}}$. In the progress of a sound-wave the air is alternately compressed and rarefied, the compressions and rarefactions occurring in such rapid succession that there is no time for any transfer of heat. If this equation be applied to air, the E becomes E_h, or the elasticity under the condition that no heat enters or escapes. Since we know the density of the air and the velocity of sound, E_h can be computed. In § 105 it is shown that E_t is numerically equal to the pressure; hence we have the values of the two elasticities of air, and, as seen above, their ratio is the ratio of the two specific heats of air.

217. Examples of Energy absorbed by Vaporization.—When a liquid boils, its temperature remains constant, however intense the source of heat. This shows that the heat applied to it is expended in producing the change of state. Heat is absorbed during evapora-

tion. By promoting evaporation, intense cold may be produced. In a vacuum, water may be frozen by its own evaporation. If a liquid be heated to a temperature above its ordinary boiling-point under pressure, relief of the pressure is followed by a very rapid evolution of vapor and a rapid cooling of the liquid. Liquid nitrous oxide at a temperature of zero is still far above its boiling-point, and its vapor exerts a pressure of about thirty atmospheres. If the liquid be drawn off into an open vessel, it at first boils with extreme violence, but is soon cooled to its boiling-point for the atmospheric pressure, about − 88°, and then boils away slowly, while its temperature remains at that low point. By liquefying nitrogen and then allowing it to evaporate under low pressure, Olszewski obtained the temperature of − 220° C., and by allowing liquid hydrogen to boil under atmospheric pressure, − 243.5° C. was reached.

218. Heat Equivalent of Vaporization.—It is plain that the formation of vapor is work requiring the expenditure of energy for its accomplishment. Each molecule that is shot off into space obtains the motion which projected it beyond the reach of the molecular attraction, at the expense of the energy of the molecules that remain behind. A quantity of heat disappears when a liquid evaporates; and experiment demonstrates, that to evaporate a kilogram of a liquid at a given temperature always requires the same amount of heat. This is the *heat equivalent of vaporization*. When a vapor condenses into the liquid state, the same amount of heat is generated as disappears when the liquid assumes the state of vapor. The heat equivalent of vaporization is determined by passing the vapor at a known temperature into a calorimeter, there condensing it into the liquid state, and noting the rise of temperature in the calorimeter. This, it will be seen, is essentially the method of mixtures. Many experimenters have given attention to this determination; but here, again, the best experiments are those of Regnault. He determined what he called the *total heat of steam* at various pressures. By this was meant the heat required to raise the temperature of a kilogram of water from zero to the temperature

of saturated vapor at the pressure chosen, and then convert it wholly into steam. The result of his experiments give, for the heat equivalent of vaporization of water at 100°, 537 calories. That is, he found that by condensing a kilogram of steam at 100° into water, and then cooling the water to zero, 637 calories were obtained. But almost exactly 100 calories are derived from the water cooling from 100° to zero; hence 537 calories is the heat equivalent of vaporization at 100°.

219. Dissociation.—It has already been noted (§ 157), that, at high temperatures, compounds are separated into their elements. To effect this separation, the powerful forces of chemical affinity must be overcome, and a considerable amount of energy must be consumed.

220. Heat Equivalent of Dissociation and Chemical Union.—From the principle of the conservation of energy, it may be assumed that the energy required for dissociation is the same as that developed by the reunion of the elements. The heat equivalent of chemical union is not easy to determine, because the process is usually complicated by changes of physical state. We may cause the union of carbon and oxygen in a calorimeter, and, bringing the products of combustion to the temperature of the elements before the union, measure the heat given to the instrument; but the carbon has changed its state from a solid to a gas, and some of the chemical energy must have been consumed in that process. The heat measured is the *available* heat. The best determinations of the available heat of chemical union have been made by Andrews, Favre and Silbermann, and Berthelot.

THE KINETIC THEORY OF HEAT.

221. Molecular Motion. States of Matter.—The continued production of heat by the expenditure of mechanical work proves that heat is not a substance, and suggests that it must be in some way dependent on motion. It has been seen that such phenomena as expansion and fusion may be explained on the hypothesis that the molecules of a body move more rapidly when the body is heated.

The emission of light or, in general, of radiant energy from a body affords a demonstration of the existence of some motion in those parts of a body which are so small that the motion cannot be directly perceived by ordinary observation; for we can explain radiance only as a motion in a medium through which it travels, and it is evident that this motion cannot be due to the mere presence of a substance, but must be set up by the motion of matter.

We may first apply the kinetic theory to the distinction between solids, liquids, and gases. Each molecule of a solid is supposed to be retained within a certain small region by the action of the surrounding molecules and to move within that region. The phenomenon of crystallization leads us to think that molecules in a solid have certain determinate forms and an arrangement in the body; their motions, therefore, are such that they do not overstep the limits of this arrangement, and we think of their motion as vibratory, using the word vibratory in a rather loose sense. The molecules of a liquid have no fixed position in the mass, but are free to move from one point to another; they are in very close proximity to one another, as appears from the phenomena of capillarity, and exert considerable forces on one another. The chief difference between solids and liquids consists in the absence in the latter of any definite arrangement; we may think of the molecules of a liquid as rotating and as gliding past each other, and can characterize their motion as rotatory. The great increase in volume exhibited on the change of a mass of liquid into vapor shows that the molecules of a vapor or gas are farther apart than those of a liquid. They are so far apart that their mutual actions due to molecular forces have very little influence on their motions, except during the excessively short period within which any two of them come close together or undergo an encounter. A molecule of a gas is therefore thought of as moving in a succession of short rectilinear paths, the direction of which is in general changed at each encounter. We may therefore characterize the motion in a gas as translatory. The consideration of this translatory motion is suffi-

cient to explain most of the laws of gases, though to explain others a rotation or something equivalent to it must be assumed.

The characteristics of the molecular motion assumed in the kinetic theory may be best explained by considering the motion in a gas. Let us suppose that a very large number of material particles is distributed uniformly throughout the region contained within a closed vessel, and that velocities are given to these molecules at a certain instant in various directions. If we further suppose that these molecules act on each other only by collision or by forces which are effective only when two molecules are extremely near each other, it is plain that the paths of the molecules thus assumed will in general be short straight lines, changing in direction with every encounter between two molecules. It is also evident that, no matter what the initial velocities were, they will not be maintained for any length of time, but that the velocity of any one molecule will change at each encounter, and that the velocities of the molecules in the mass will speedily acquire values ranging from zero to a very great or practically infinite velocity. It is also plain that very few molecules will possess these extreme velocities at any one time, and that most of them will possess velocities which do not depart far from a certain mean. An obvious condition to which the velocities must conform is that the kinetic energy of all the molecules in the mass must remain the same at all times, it being assumed that no energy enters the mass from without and that the encounters do not involve the loss of kinetic energy. It was shown by Clausius, and afterwards more rigorously by Maxwell, that the distribution of velocity among the molecules may be deduced by the theory of probabilities. Some idea of it may be got from the distribution of shots in a target; if a rifleman shoot at a target a great many times, and if the distance of the shots from the centre of the bull's eye be measured, these distances conform to the same law of distribution. It is clearly infinitely improbable that any one of the shots will strike the exact centre of the bull's eye, and also infinitely improbable that any one will be sent directly away from the target, and it is very highly improb-

able that any one will miss the target entirely; the vast majority of the shots will meet the target, and their distances from the centre will lie around a certain average distance. Similarly, it is extremely improbable that any molecule of a gas will have a velocity far exceeding the average; the great majority of them will have velocities which lie around a certain mean velocity. The law of distribution of velocities among molecules of liquids and solids is not known, but it probably possesses the essential characteristics of the law for gases.

When a gas is heated, all but a very small part of the heat which enters it is used in increasing the kinetic energy of the molecules; this is not true for solids and liquids, because, when they are heated, work is done against their molecular forces which does not appear as kinetic energy. The kinetic energy of the molecule is the sum of the kinetic energy due to the motion of its centre of mass or to its translation, and of the kinetic energy due to its motion relative to its centre of mass. This latter energy may be thought of as due either to rotation about the centre of mass or to the vibrations of the atoms constituting the molecule. We will subsequently prove that the temperature of a gas is proportional to the kinetic energy of its molecules. It is therefore natural to assume that the measure of temperature is some part of the kinetic energy of the molecule. The most consistent explanation of all the effects of heat can be reached by supposing that the energy of atomic vibration or of molecular rotation is directly proportional to the temperature measured on the absolute scale (§ 212). The total kinetic energy of the molecules of a body measures the heat in the body.

222. Kinetic Theory of Gases.—The foundation of the theory of matter now under discussion is the *kinetic theory of gases.* In this theory a perfect gas consists of an assemblage of free, perfectly elastic molecules in constant motion. Each molecule moves in a straight line with a constant velocity, until it encounters some other molecule, or the side of the vessel. The impacts of the molecules

upon the sides of the vessel are so numerous that their effect is that of a continuous constant force or pressure.

The entire independence of the molecules is assumed from the fact that, when gases or vapors are mixed, the pressure of one is added to that of the others; that is, the pressure of the mixture is the sum of the pressures of the separate gases. It follows from this, that no energy is required to separate the molecules; in other words, no internal work need be done to expand a gas. This was demonstrated experimentally by Joule (§ 215).

The action between two molecules, or between a molecule and a solid wall, must be of such a nature that no energy is lost; that is, the sum of the kinetic energies of all the molecules must remain constant. Whatever be the nature of this action, it is evident that when a molecule strikes a solid stationary wall it must be reflected back with a velocity equal to that before impact. If the velocity be resolved into two components, one parallel to the wall and the other normal to it, the parallel component remains unchanged, while the normal component is changed from $+u$, its value before impact, to $-u$, its value after impact. The change of velocity is therefore $2u$, and if θ represent the duration of impact, the mean acceleration is $\frac{2u}{\theta}$, and the mean force of impact $p = m\frac{2u}{\theta}$, where m represents the mass of the molecule.

Since the effect of the impacts is a continuous pressure, the total pressure exerted upon unit area is equal to this mean force of impact of one molecule multiplied by the number of molecules meeting unit area in the time θ. To find this latter factor, we suppose the molecules confined between two parallel walls at a distance s from each other. Any molecule may be supposed to suffer reflection from one wall, pass across to the other, be reflected back to the first, and so on. Whatever may be the effect of the mutual collisions of the molecules, the number of impacts upon the surface considered will be the same as though each one preserved its rectilinear motion unchanged, except when reflected from the solid walls. The time required for a molecule moving with a velocity u

to pass across the space between the two walls and back is $\frac{2s}{u}$; and the number of impacts upon the first surface in unit time is $\frac{u}{2s}$.

Consider the molecules contained in a rectangular prism, with bases of area a in the walls. These molecules must be considered as moving in all directions and with various velocities. But the velocity of any molecule may be resolved in the direction of three rectangular axes, one normal to the surface and the other two parallel to it, and the effect upon the walls will be due only to the normal components. Let us single out for examination a group of molecules which have a normal velocity that lies near the value u_1, and let n_1 represent the number of such molecules in unit volume. Then the number of such molecules within the prism considered is n_1sa. The number of impacts made by them in unit time on one of the walls is $n_1sa \cdot \frac{u_1}{2s} = \frac{n_1au_1}{2}$, and in the time θ is $\frac{n_1au_1\theta}{2}$. Hence the total pressure which they exert on the area a is $m\frac{2u_1}{\theta} \cdot \frac{n_1au_1\theta}{2} = mn_1u_1^2a$, and on unit area is $mn_1u_1^2$.

Now the total pressure on unit of area is the sum of the pressures due to all the i groups into which the molecules of the gas may be divided, or $p = m(n_1u_1^2 + n_2u_2^2 + \ldots n_iu_i^2)$. If we represent by n the number of molecules in unit volume and by u the mean velocity given by $nu^2 = n_1u_1^2 + n_2u_2^2 + \ldots n_iu_i^2$, we have $p = mnu^2$. Similar expressions hold for the pressures on the other walls, the velocities normal to which are v and w, and we assume that these mean velocities are independent of direction, so that $u^2 = v^2 = w^2$. But the velocity of any molecule is given by $V_i^2 = u_i^2 + v_i^2 + w_i^2$, and the mean velocity is given by a similar equation. Hence $V^2 = 3u^2$, and we have finally,

$$p = \frac{1}{3}mnV^2. \tag{74}$$

The velocity V in this expression is called the *velocity of mean square*.

If we now suppose the volume of the gas to change so that the

volume which contains n molecules becomes v, the pressure takes a new value, which we will still designate by p. We have

$$p = \frac{1}{3} m \frac{n}{v} V^2, \quad \text{or} \quad pv = \frac{1}{3} mn V^2. \tag{75}$$

Since V^2 remains constant so long as the temperature is constant, and since m and n are fixed, we have pv constant. Hence Boyle's Law follows from the kinetic theory of gases.

From Gay-Lussac's law (§ 211) it has been shown that if we reckon temperature from $-273°$ C. as a zero, we have $pv = RT$ for all gases. Using the equation just proved, we have

$$RT = \frac{1}{3} mn V^2. \tag{76}$$

Now $\frac{1}{2} m V^2$ is the mean kinetic energy of the molecule. The formula shows, therefore, that the temperature on the scale of the air-thermometer is proportional to the mean kinetic energy of the molecule. The zero of this scale will be that temperature at which $V = 0$, or at which the molecules are at rest. There can be no temperature lower than this, and hence we obtain a warrant for calling this temperature a real or *absolute zero*. The final demonstration of the existence of such a zero will be given in § 231, where it is not based upon any particular theory of matter.

It was demonstrated by Maxwell that the mean kinetic energies of the molecules of different gases at the same temperature are the same, or that $\frac{1}{2} m_1 V_1^2 = \frac{1}{2} m_2 V_2^2$. If we consider equal volumes of two gases at the same pressure and temperature, for which, therefore, $\frac{1}{3} m_1 n_1 V_1^2 = \frac{1}{3} m_2 n_2 V_2^2$, we obtain $n_1 = n_2$, or the numbers of molecules of the two gases in the same volume are the same. This is *Avogadro's law*.

Up to this point we have considered the molecules as particles, and have supposed that all their energy exists as the kinetic energy of molecular motion. It is easy to show, however, that this supposition is in error, and that the molecules possess more energy than that given by $\frac{1}{2} mn V^2$. Let us consider a unit mass of gas, the temperature of which is raised under constant pressure by a small amount ΔT. Then the work done by its expansion (§ 215) is rep-

resented by $(C_p - C_v)\varDelta T$. We shall show (§ 232) that this work is also given by the product of the pressure by the increase in volume, or by $p\varDelta v$. Hence we have the relation $p\varDelta v = (C_p - C_v)\varDelta T$.

The kinetic energy of molecular translation is $\frac{1}{2}mnV^2$, and if C_0 represent its increase for a rise in temperature of one degree, $C_0\varDelta T$ represents its increase for the rise of temperature $\varDelta T$. But since $pv = \frac{1}{3}mnV^2$, we have $C_0\varDelta T = \frac{3}{2}p\varDelta v$, and hence $C_p - C_v = \frac{2}{3}C_0$, or

$$\frac{C_0}{C_v} = \frac{3}{2}\left(\frac{C_p}{C_v} - 1\right). \tag{77}$$

Now on the supposition that the molecules are particles which have no energy except energy of translation, $C_0 = C_v$, and hence $\frac{C_p}{C_v} = \frac{5}{3}$. We know by experiment that this is not always the case. For monatomic gases, such as mercury vapor, and possibly argon, $\frac{C_p}{C_v} = 1.66$; but for the common diatomic gases it is more nearly $\frac{7}{5} = 1.4$, and for gases with complex molecules, it is about $\frac{4}{3} = 1.33$. Hence, in the case of gases with more than one atom in the molecule, the total energy is not merely the energy of translation, but includes other energy internal to the molecule. Boltzmann has shown that the ratio of the internal energy to the energy of translation is such as can be accounted for by supposing the monatomic molecules to be spheres or points, the diatomic molecules solids of revolution, and the more complex molecules irregular solids. It is likely that this is merely an artificial representation, since there is strong reason to believe that the atoms vibrate within the molecule and that the molecule is not rigid.

We have used C_0 to represent the increase in the energy of molecular translation in a unit of mass when the temperature rises one degree. If we represent the increase in the kinetic energy of a single molecule by $\varDelta\frac{1}{2}mV^2$, we have $C_0 = n\varDelta\frac{1}{2}mV^2$. Now n is the number of molecules in unit volume, which in this equation is the volume containing unit mass, so that $\frac{1}{n}$ is the mass of one molecule or m. The gain in kinetic energy for a rise of temperature of

one degree is, by Maxwell's law, the same for all gases, so that $\Delta\frac{1}{2}mV^2$ is a constant for all gases; and hence C_0m is a constant for all gases. C_0 cannot be directly observed, but we may set $\frac{C_v}{C_0} = \beta$, and observe C_v. If β is the same for all gases, C_vm should also be constant. This is *Dulong and Petit's law* for gases. It holds quite closely for all gases of the same type of molecular structure, and the departures from it are readily explained by the probability that β is not the same for all gases.

The phenomena exhibited by the *radiometer* afford a strong experimental confirmation of the kinetic theory of gases. These phenomena were discovered by Crookes. In the form first given to it by him, the instrument consists of a delicate torsion balance suspended in a vessel from which the air is very completely exhausted. On one end of the arm of the torsion balance is fixed a light vane, one face of which is blackened. When a beam of light falls on the vane it moves as if a pressure were applied to its blackened surface. The explanation of this movement is, that the molecules of air remaining in the vessel are more heated when they come in contact with the blackened face of the vane than when they come in contact with the other face, and are hence thrown off with a greater velocity, and react more strongly upon the blackened face of the vane. At ordinary pressures the free paths of the molecules are very small, their collisions very frequent, and any inequality in the pressures is so speedily reduced that no effect upon the vane is apparent. At the high exhaustions at which the movement of the vane becomes evident, the collisions are less frequent, and hence an immediate equalization of pressure does not occur. The vane therefore, moves in consequence of the greater reaction upon its blackened surface.

223. Influence of the Size and Attractions of the Molecules. Critical Temperature.—On the elementary theory which has just been developed all gases should conform precisely to the so-called gaseous laws, whereas in fact they only conform to those laws approximately. It was shown by van der Waals that the devia-

tions exhibited by gases from the gaseous laws can be accounted for by extending the theory so as to include the consideration of the size of the molecules and of their mutual attractions. In the elementary theory the molecules were assumed to be points or particles of negligible magnitude, but if we assume them to have volumes which, though small, are appreciable, it is plain that the effective volume within which the molecules have free motion is reduced by an amount dependent on the molecular volumes. It was furthermore assumed in the elementary theory that the time of encounter is negligible in comparison with the time during which the molecule is free from the action of other molecules; but if we assume that the time of an encounter, though small, is not negligible, it is plain that the molecular attractions will tend to hold together the mass of gas or will be equivalent to an addition to the pressure upon the gas. From these considerations van der Waals expressed the relations among the pressure, volume, and temperature of a gas by the formula

$$\left(p + \frac{a}{v^2}\right)(v - b) = RT, \tag{78}$$

where a is a constant depending upon the molecular attractions, and b is four times the sum of the volumes of the molecules. This formula, when tested by experiment, represents the behavior of gases far more accurately than the simpler form; it is not, however, exact, and various others, constructed empirically, have been proposed which give even a better representation of the facts. It is as yet the only formula for which a theoretical demonstration has been given. This formula possesses the great advantage that it can represent the behavior of a body, at least in certain cases, not only in the gaseous but in the liquid state; that is, it exhibits the continuity which we have every reason to think exists between those states. In particular it gives an explanation of critical temperature and a determination of it in terms of the molecular constants a and b. If the formula be expanded and arranged in the order of the descending powers of v, it becomes $v^3 - v^2\left(b + \frac{RT}{p}\right) + v\frac{a}{p} - \frac{ab}{p} = 0.$

This is a cubic equation, and, for given values of p and T, will have three roots, which are either all real or of which one is real and the others imaginary, depending upon the values of the constants and on the particular values chosen for p and T. The existence of three real roots shows that for the assumed values of pressure and temperature three different volumes are possible; one of these is the volume of the body as a gas, another its volume as a liquid, and the third its volume in an intermediate or transition state which is unstable. The existence of only one real root shows that, for the particular values of pressure and temperature which give it, the substance can exist in only one state, either as a liquid or as a gas. The study of the real roots shows that, as the pressure and temperature increase, the values of the roots become more nearly equal, until for a certain definite pressure and temperature they become coincident; when the value of the temperature is still higher two of the roots cease to be real. The temperature which corresponds to the existence of three coincident roots is the *critical temperature;* at any temperature higher than that the substance can exist only as a gas.

224. General Explanation of Liquefaction, etc.—We will now apply the kinetic theory to give an explanation of the principal phenomena exhibited by a substance as it is heated. Let us consider a substance in the solid state at a temperature below its melting point; suppose heat applied to it gradually and at a uniform rate from some source. Its temperature will rise and it will in general expand; the rise of temperature is of course due to the increase in that part of the kinetic energy of the body which is the measure of temperature; on the view we have adopted, to the increase in the kinetic energy of molecular rotation or atomic vibration. The expansion is explained by the increase in the kinetic energy of translation, which enables the molecules to move farther from one another and so to increase the regions occupied by them. When the temperature rises to the melting-point, these regions have become so large that the molecules in them are no longer constrained to any definite directions; their motions there-

fore become rotatory and they are free to glide past each other in the mass. We can explain the constancy of temperature during melting, and the absorption of heat, by assuming that that portion of the energy which measures temperature remains constant, and that the heat is used in doing work against the molecular forces which determine the direction of the molecules in the solid and in giving the molecules increased velocity of translation. Such a change as is here described, in which the energy received by the molecule does work against the forces acting on it and gives it greater velocity as a whole, while the mean energy of vibration which it had at first is equal to the mean energy of rotation which it acquires, has been shown by Eddy to be mechanically possible. On melting, the body generally changes its volume, sometimes expanding, sometimes contracting. This may be explained by supposing that as the molecules are heated, their volumes diminish. The admissibility of this assumption has been proved by Lorentz and Sutherland. The change in volume on melting is then the resultant of the expansion due to the increased molecular motion and the contraction due to the shrinking of the molecules, and it may therefore be either positive or negative.

After melting, the temperature of the body continues to rise and the body generally expands until the boiling-point is reached; at that point the temperature again ceases to rise and the liquid becomes a vapor. We explain this by supposing that in consequence of the changes in velocity which go on among the molecules, there will arise an assemblage of molecules in a small region with velocities above the average; these will beat back the surrounding molecules and form a small bubble within which the molecules are in the gaseous state. Those molecules near the surface of this bubble which possess velocities above the average will pass through the liquid surface against the attractions of the molecules surrounding them and will increase the gas contained in the bubble, until its size becomes such that its buoyancy is able to overcome the viscosity of the liquid, so that it rises and sets free a number of molecules at the surface of the liquid in the gaseous state. The equality of

temperature between the liquid and the vapor formed from it, and the absorption of heat during this process, are explained by supposing that that part of the kinetic energy which measures temperature remains constant, and that the heat is used in doing work against the molecular forces which determine the volume of the liquid. Any further heating of the vapor increases its total kinetic energy and that part of it which measures temperature in nearly the same proportion.

The specific heat of the substance increases when it passes from the solid to the liquid state, and decreases when it becomes a gas. This is explained by the supposition, which many facts render probable, that the kinetic energy of translation of the molecule is greater in the liquid state than in either of the other two states in comparison with the kinetic energy of rotation or of atomic vibration.

The explanation of evaporation which goes on from many solids and liquids at all temperatures has been already given (§ 202); it depends upon the fact that the velocity of some of the molecules is always far greater than the average velocity, and may be sufficient to carry those molecules beyond the range of molecular action.

The hypothesis that the temperature is measured by the kinetic energy of rotation or of atomic vibration is confirmed by its application to Dulong and Petit's law; as it is our purpose to give a general idea of the theory rather than a defence of it, we will not enter upon the discussion of this point.

225. Molecular Velocities and Dimensions. — The formula $pv = \frac{1}{2} mn V^2$ enables us to calculate V, the velocity of mean square, since mn is the mass in the volume v, and p can be measured in absolute units. If we apply the equation to hydrogen under atmospheric pressure, we have $p = 1013373$ dynes per square centimetre and $\frac{mn}{v}$, or the density, $= 0.00008954$ grams per cubic centimetre, and hence $V = 184260$ centimetres per second, or a little more than a mile per second. Since for different gases with the same pressure, volume, and temperature V^2 is inversely as m, the velocities in the

other gases can be found by dividing this velocity by the square root of the ratio of their masses to the mass of the molecule of hydrogen, or by the square roots of their molecular weights divided by 2.

From calculations based on the behavior of gases with reference to their viscosity and thermal conductivity, Maxwell deduced a number of conclusions respecting the dimensions and motions of molecules, which are given in the following table. The symbol $\mu\mu$ denotes a micromillimetre, or the millionth of a millimetre.

	Hydrogen.	Oxygen.	Carbon dioxide.
Mean free path in $\mu\mu$	96.5	56	43
Number of collisions per sec.	$1.775 \cdot 10^{10}$	$0.7646 \cdot 10^{10}$	$0.972 \cdot 10^{10}$
Diameter in $\mu\mu$, molecules supposed spherical	0.58	0.76	0.93
Mass in 10^{-26} grams	46	736	1012

The number of hydrogen molecules in a milligram is about 200 million million million, and about 2 million could be placed side by side in one millimetre. The number of molecules of hydrogen, and so also of any other gas, in one cubic centimetre at the standard pressure and temperature is about 19 million million million.

From the experiments of Quincke and Reinhold and Rücker the range of molecular action is estimated to lie between 50 $\mu\mu$ and 118 $\mu\mu$. The molecular forces give rise to pressures in the gas which van der Waals estimates as, for hydrogen, 0; for air, 0.0028; for carbon dioxide, 0.00874.

Other calculations yield values for these various molecular constants which, while not numerically the same as those of Maxwell, are yet of the same order of magnitude, and considerable confidence can be placed in their general accuracy.

CHAPTER IV.

THERMODYNAMICS.

226. First Law of Thermodynamics.—The law of the conservation of energy, in the special case of heat and mechanical work, is called the *first law of thermodynamics*. It may be thus stated: When heat is transformed into work, or work into heat, the quantity of work is equivalent to the quantity of heat. The experiments of Joule, Rowland, and Hirn establishing this law, and determining the mechanical equivalent, have already been described (§ 171).

227. The Thermodynamic Engine.—When a body does work against non-conservative forces, so that heat is evolved, the operations may be so regulated that all, or practically all, of the work done is transformed into heat. On the other hand, if a certain quantity of heat be present in a body, from which it may be drawn in any manner, so that it can be used for the doing of work, it is never possible, under conditions attainable on the earth's surface, even if they were ideally perfect, to transform the whole of this heat into work. The operations necessary for the transformation of some of it involve the transfer of the rest to other bodies of lower temperature.

The operation of transforming heat into work is in general very complicated; it is, however, possible to conceive of a simple operation by means of which heat may be transformed into work, and in which a relation may be found between the quantities of heat and the temperatures concerned. The relations thus developed may then be extended to far more complicated cases.

An arrangement designed to transform heat into work is called

an *engine.* In the ideal form it consists of a body called the *source*, from which heat may be drawn, another body called the *refrigerator*, into which heat may be sent, and a third body called the *working body,* which expands or contracts on the reception or emission of heat. The working body will itself always possess energy in the form of heat and possibly also in other forms. If heat be supplied to it from the source, it will expand and do work, but no relation can be stated between the work done and the heat supplied to it, because the change in its own energy experienced during the expansion is, in general, unknown. In order to obtain a relation between the heat supplied to the working body and the work done by it, the operations performed with it must be so conducted as to bring the working body back to its original state. It will then possess the same energy as at the outset, and the first law of thermodynamics enables us to assert that the difference between the heat which leaves the source and the heat which enters the refrigerator is equal to the work done by the working body. Such a series of operations is called a *cycle.* The ratio of the work done to the heat which leaves the source is called the *efficiency* of the engine.

228. The Carnot's Cycle.—In order to study the efficiency of an engine we restrict the conditions under which the transformation of heat into work goes on. We suppose that the source is so large and furnishes so unlimited a supply of heat that its temperature S remains constant, notwithstanding the loss or gain of heat which it may receive from the working body. Similarly, we suppose the refrigerator to have a constant temperature R, notwithstanding the gain or loss of heat it may receive from the working body. The changes by which the working body does work are supposed to occur only when the working body is either at the temperature of the source or of the refrigerator, or when it is so conditioned that it neither receives nor emits heat. While it is kept at a constant temperature, its change is isothermal; when it neither receives nor emits heat, its change is adiabatic (§ 213).

In order to exhibit the operation of this simple engine most clearly, we will assume that the working body is one which

increases in volume on the introduction of heat, and does work by expansion, under a pressure which is the same for all points on its surface.

Construct two rectangular axes of volume and pressure as in Fig. 72. Let the pressure and corresponding volume of the body, when its temperature R is that of the refrigerator, be represented by the point A. The operation of transforming the heat received from the source into work is then performed as follows: The body, being so placed that it can neither receive nor emit heat, is compressed adiabatically;

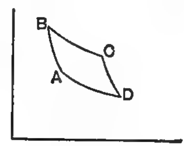

FIG. 72.

its volume diminishes and its temperature rises, and the operation is continued until its temperature becomes S, equal to that of the source. Its pressure and volume in this state are represented by the point B. It is then put in contact with the source and allowed to expand; as soon as its expansion begins, its temperature falls and heat enters from the source. The expansion may be so regulated that the difference of temperature between the body and the source never exceeds an infinitesimal, so that the heat which enters the body during this part of the process enters it at the constant temperature S. The expansion may be allowed to continue until any desired quantity of heat H is taken from the source. The pressure and corresponding volume attained by this isothermal expansion are represented by the point C. The body is then removed from the source and allowed to continue its expansion under such conditions that it neither receives nor emits heat. Its volume will increase and its temperature will fall. This adiabatic expansion is allowed to continue until the temperature of the body becomes R, that of the refrigerator. The body is then placed in contact with the refrigerator and compressed. As its volume begins to diminish, its temperature rises, and heat passes out from it into the refrigerator. The compression may be so regulated that the difference of temperature between the body and the refrigerator never exceeds an infinitesimal, so that the heat which leaves the body during this part of the process leaves it at the constant tem-

perature R. The operation is continued until the volume and pressure of the body are again denoted by the point A. During this operation the quantity of heat h is transferred from the body to the refrigerator. These operations constitute a cycle, for the body at the end of the operation is in the same condition as regards pressure, volume, and temperature as it was at the beginning. The work done by it is therefore equal to the heat transformed into work, or to $H-h$.

Such a cycle is *reversible*, for if the body be constrained to go through the operations just described, in the reverse order, the same quantities of heat will be transferred in opposite senses and the same quantity of work done upon the body that, in the direct operation, was done by the body. That is, the refrigerator will give up the quantity of heat h, the source will receive the quantity of heat H, and the amount of work $H-h$ will be done upon the body. The only difference between the two operations will be that, whereas in the direct operation the temperature of the body was infinitesimally lower than that of the source while it was receiving heat, and infinitesimally higher than that of the refrigerator while it was emitting heat, in the reversed operation the temperature of the body is infinitesimally lower than that of the refrigerator while it is receiving heat, and infinitesimally higher than that of the source while it is emitting heat. These infinitesimal differences may be neglected, and one of these operations may be considered in every respect the reverse of the other.

229. Second Law of Thermodynamics.—We will now prove a most important proposition, due to Carnot, the founder of the theory of thermodynamics. To do this we make use of a principle first laid down by Clausius and known either as *Clausius's principle* or the *second law of thermodynamics*. This principle is, that heat cannot pass of itself, or without compensation in the form of work done or of heat transferred in the opposite sense, from a colder to a hotter body. This principle is in conformity with our common experience, that heat passes by conduction or radiation from a place of higher to a place of lower temperature. It is not

susceptible of immediate demonstration, and is accepted as a general principle for reasons similar to those which determine the acceptance of Newton's laws of motion as statements of general truths respecting motion.

230. Efficiency of a Reversible Engine.—Carnot's proposition, which is now to be proved, asserts that the efficiency of all reversible engines is the same. To show this, let us suppose a reversible engine A and a non-reversible engine B, working between the same source and the same refrigerator, and let us assume that the efficiency of the non-reversible engine B is greater than that of the reversible engine A. Let the engine B work forward, so as to do work W and give to the refrigerator the quantity of heat h_B; it will therefore take from the source the quantity of heat $H_B = W + h_B$. Let the work W be expended in driving the reversible engine A backward. The engine A will take from the refrigerator the quantity of heat h_A and give to the source the quantity of heat $H_A = W + h_A$. Now, by hypothesis, the efficiency of the non-reversible engine B is the greater, so that $\frac{W}{H_B} > \frac{W}{H_A}$, and therefore $H_A > H_B$, and also $h_A > h_B$. The result of these combined operations is that no work is done by the engines, and that the source receives heat while the refrigerator loses heat. This conclusion is contrary to Clausius's principle and must be rejected, as inconsistent with the operations of Nature. We conclude, therefore, that no engine can have an efficiency greater than that of the reversible engine. It follows as a corollary that the efficiency of all reversible engines is the same.

231. Absolute Scale of Temperatures.—Since the efficiency of all reversible engines is the same and is a maximum, it is manifestly indifferent what material is used in the working body; indeed, since the demonstration just given does not involve as essential the particular mode of doing work assumed in the construction of the diagram, it is also indifferent in what way the working body changes its dimensions and does work. The work done depends only on the heat received from the source and on the

temperature of the source and refrigerator, and the efficiency depends only on the temperatures of the source and the refrigerator, or is a function of these temperatures. If the temperatures be represented on any conventional scale, the form of this function may be found by experiment; on the other hand, the assumption of a form of this function will determine a scale of temperatures. The proposal to form such a scale, which is dependent only on the efficiency of the reversible engine, and is therefore independent of the properties of any particular body, was made by William Thomson.

The scale of temperatures which is most convenient for application in thermodynamics, and which is so distinguished by its simplicity from all others that might be formed that it is called distinctively *the absolute scale* of temperatures, is formed by assuming that the efficiency of a reversible engine is equal to the ratio of the difference of temperature between the source and the refrigerator and the temperature of the source, that is, by assuming

$$\frac{H-h}{H}=\frac{S-R}{S}. \tag{79}$$

This assumption may also be stated in the form

$$\frac{h}{H}=\frac{R}{S}. \tag{80}$$

The maximum efficiency of an engine is attained when all the heat which is received from the source is transformed into work, so that no heat is transferred to the refrigerator; on the scale of temperatures just assumed this condition is attained when $R=0$. This zero is an absolute and not an arbitrary zero. It depends on the general properties of bodies, and not on the particular properties of any one body. It is the lowest temperature attainable in Nature, for, if it were possible to have a refrigerator at a lower temperature than this, the efficiency of an engine working with that temperature as the temperature of its refrigerator, would be

greater than unity. This temperature is therefore called the *absolute zero.*

The length of the degree on the absolute scale may be determined by designating the difference of temperature between two bodies by an arbitrarily chosen number and by measuring the efficiency of an engine working between the temperatures of those bodies. The most convenient assumption to make is that the absolute difference between the temperature of boiling water and the temperature of melting ice is 100 degrees. The temperature intervals or degrees on the scale thus formed are very nearly those of the Centigrade scale.

232. Relation of the Absolute Temperature to the Temperature of the Air Thermometer.—Let us assume that a substance exists which obeys perfectly the laws of Boyle and Gay-Lussac; such a substance is called a *perfect gas.* We wish to show that the temperatures indicated by the expansion of a perfect gas, used as a thermometric substance, will be those of the absolute scale.

We must first prove that the work done by the expansion of a gas is equal to the area included between the lines representing its changes of pressure and volume, the two ordinates representing its extreme pressures and the horizontal line of zero volume. The proof of this proposition does not depend on the properties of a perfect gas, and the proposition holds in all cases in which the body does work by expanding under a hydrostatic pressure which is the same at all points of its surface. Let us select a small area s on the surface of the body. The pressure p is applied to all points of the surface, and the force which acts on the area s is therefore ps. Let the body expand slightly, so that the area s is displaced along its normal through the distance n. The work done in displacing the area s is psn, and the work done in expanding the whole body is $\Sigma psn = p\Sigma sn$, since p is the same for all points on the surface. Now Σsn is equal to the increase in the volume of the body, or to dv. The work done during the small expansion is therefore pdv. This expansion will, in general, involve an infinitesimal change in the pressure; but if the process here described be repeated for each

infinitesimal increment of volume, the sum of all the terms pdv will equal the total work done by the expansion of the body. Now let us consider the area $bBCc$ standing under the line BC (Fig. 73). This area may be conceived of as made up of a series of infinitesimal rectangles, the heights of which are the ordinates of the successive points of the line BC, and the bases of which are successive elements taken along the line bc. If dv represent the length of one of these elements, and p the corresponding ordinate, the area of the infinitesimal rectangle determined by them is pdv. The sum of such areas for the expansion indicated by the line BC is the area $bBCc$; and since Σpdv represents the work done, the area $bBCc$ also represents the work done during the expansion of the body in the way indicated by the line BC.

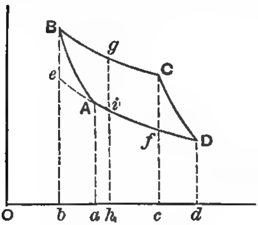

FIG. 73.

Now to demonstrate the relation between the temperatures indicated by the perfect gas thermometer and those of the absolute scale, let us suppose an engine in which the working body is a perfect gas, and let us suppose that the changes in pressure and volume experienced by the working body during the cycle are so small that the portions of the isothermal and adiabatic lines which bound it are straight, and that the cycle is a parallelogram. This cycle is represented by the area $ABCD$ (Fig. 73). We may assume as the result of the experiments of Joule that when a gas expands at constant temperature, no internal work is done upon it, or that the heat which enters it is entirely spent in doing external work. Produce DA to e; then the parallelogram $ABCD$ is equal to the parallelogram $eBCf$, and this parallelogram represents the work done in the cycle by the gas acting as the working body.

The work done during the expansion from B to C, which is equal to the heat received during that expansion, is represented by the area $bBCc$. Let g be the middle point of the line BC; the perpendicular gh will bisect the line ef at i. The area $bBCc = bc \cdot gh$, and the area $eBCf = bc \cdot gi$. Therefore the efficiency of

the engine, or $\frac{eBCf}{bBCc}$, equals $\frac{gi}{gh}$. Now gh represents the pressure of the gas at the temperature t of the source, when its volume is Oh, and gi represents the diminution of pressure caused by a fall of temperature to θ, the temperature of the refrigerator, when the volume is kept constant. The efficiency of the engine is therefore $\frac{p_t - p_\theta}{p_t}$. And since the efficiency is also given by $\frac{S - R}{S}$, where S and R are the temperatures of source and refrigerator on the absolute scale, $\frac{S - R}{S} = \frac{p_t - p_\theta}{p_t}$ or $\frac{R}{S} = \frac{p_\theta}{p_t}$. We know, from the experiments of Gay-Lussac, that if t and θ be measured on the Centigrade scale, and if p_0 represent the pressure of the gas at the Centigrade zero on the condition that the volume is constant, $p_t = p_0(1 + \alpha t)$ and $p_\theta = p_0(1 + \alpha\theta)$, where $\alpha = \frac{1}{273}$ is the coefficient of expansion. Using these values in the above equation we obtain $\frac{R}{S} = \frac{1 + \alpha\theta}{1 + \alpha t} = \frac{273 + \theta}{273 + t}$. If the pressure or volume of a gas, the two being interchangeable by Boyle's law, be used as a measure of its temperature, the pressure or volume and the temperature will always be directly proportional, provided the zero of temperature be taken at $-273°$ Centigrade; this temperature is the zero of the perfect gas thermometer. From the equation just obtained it is clear that the absolute scale of temperatures is the same as the one given by the perfect gas thermometer, and that the absolute zero is the zero of the perfect gas thermometer.

No gases conform precisely to the laws of Boyle and Gay-Lussac, and consequently no gas thermometer can be constructed which will accurately indicate the absolute scale of temperatures. Nevertheless, some gases depart only slightly from the conditions of a perfect gas, and the temperature determinations given by thermometers in which such gases are employed may be converted by suitable corrections into the corresponding absolute temperatures.

233. The Steam-engine.—The *steam-engine* in its usual form consists essentially of a piston, moving in a closed cylinder, which is provided with passages and valves by which steam can be admitted and allowed to escape. A boiler heated by a suitable furnace supplies the steam. The valves of the cylinder are opened and closed automatically, admitting and discharging the steam at the proper times to impart to the piston a reciprocating motion, which may be converted into a circular motion by means of suitable mechanism.

There are two classes of steam-engines, *condensing* and *non-condensing*. In condensing engines the steam, after doing its work in the cylinder, escapes into a condenser, kept cold by a circulation of cold water. Here the steam is condensed into water; and this water, with air or other contents of the condenser, is removed by a pump. In non-condensing engines the steam escapes into the open air. In this case the temperature of the refrigerator must be considered at least as high as that of saturated steam at the atmospheric pressure, or about 100°, and the temperature of the source must be taken as that of saturated steam at the boiler-pressure. Applying the expression for the efficiency (§ 231), $e = \frac{S - R}{S}$, it will be seen that, for any boiler-pressure which it is safe to employ in practice, it is not possible, even with a perfect engine, to convert into work more than about fifteen per cent of the heat used.

In the condensing engine the temperature of the refrigerator may be taken as that of saturated steam at the pressure which exists in the condenser, which is usually about 30° or 40°: hence $S - R$ is a much larger quantity for condensing than for non-condensing engines. The gain of efficiency is not, however, so great as would appear from the formula, because of the energy that must be expended to maintain the vacuum in the condenser.

234. Hot-air and Gas Engines.—Hot-air engines consist essentially of two cylinders of different capacities, with some arrangement for heating air in, or on its way to, the larger cylinder. In one form of the engine an air-tight furnace forms the passage be-

tween the two cylinders, of which the smaller may be considered as a supply-pump for taking air from outside and forcing it through the furnace into the larger cylinder, where, in consequence of its expansion by the heat, it is enabled to perform work. On the return stroke this air is expelled into the external air, still hot, but at a lower temperature than it would have been had it not expanded and performed work. This case is exactly analogous to that of the steam-engine, in which water is forced, by a piston working in a small cylinder, into a boiler, is there converted into steam, and then, acting upon a much larger piston, performs work, and is rejected. In another form of the engine, known as the "ready motor," the air is forced into the large cylinder through a passage kept supplied with crude petroleum. The air becomes saturated with the vapor, forming a combustible mixture, which is burned in the cylinder itself.

The Stirling hot-air engine and the Rider "compression-engine" are interesting as realizing an approach to Carnot's cycle.

These engines, like those described above, consist of two cylinders of different capacities, in which work air-tight pistons; but, unlike those, there are no valves communicating with the external atmosphere. Air is not taken in and rejected; but the same mass of air is alternately heated and cooled, alternately expands and contracts, moving the piston, and performing work at the expense of a portion of the heat imparted to it.

It is of interest to study a little more in detail the cycle of operations in these two forms of engines. The larger of the two cylinders is kept constantly at a high temperature by means of a furnace, while the smaller is kept cold by the circulation of water. The cylinders communicate freely with each other. The pistons are connected to cranks set on an axis, so as to make an angle of nearly ninety degrees with each other. Thus both pistons are moving for a short time in the same direction twice during the revolution of the axis. At the instant that the small piston reaches the top of its stroke, the large piston will be near the bottom of the cylinder, and descending. The small piston now descends, as well

as the large one, the air in both cylinders is compressed, and there is but little transfer from one to the other. There is, therefore, comparatively little heat given up. The large piston, reaching its lowest point, begins to ascend, while the descent of the smaller continues. The air is rapidly transferred to the larger heated cylinder, and expands while taking heat from the highly heated surface. After the small piston has reached its lowest point there is a short time during which both the pistons are rising and the air expanding, with but little transfer from one cylinder to the other, and with a relatively small absorption of heat. When the descent of the large piston begins, the small one still rising, the air is rapidly transferred to the smaller cylinder: its volume is diminished, and its heat is given up to the cold surface with which it is brought in contact. The completion of this operation brings the air back to the condition from which it started. It will be seen that there are here four operations, which, while not presenting the simplicity of the four operations of Carnot,—since the first and third are not performed without transfer of heat, and the second and fourth not without change of temperature,—still furnish an example of work done by heat through a series of changes in the working substance, which brings it back, at the end of each revolution, to the same condition as at the beginning.

Gas-engines derive their power from the force developed by the combustion, within the cylinder, of a mixture of illuminating gas and air.

As compared with steam-engines, hot-air and gas engines use the working substance at a much higher temperature. $S-R$ is, therefore, greater, and the theoretical efficiency higher. There are, however, practical difficulties connected with the lubrication of the sliding surfaces at such high temperatures that have so far prevented the use of large engines of this class.

235. Sources of Terrestrial Energy.—Water flowing from a higher to a lower level furnishes energy for driving machinery. The energy theoretically available in a given time is the weight of the water that flows during that time multiplied by the height of

the fall. If this energy be not utilized, it develops heat by friction of the water or of the material that may be transported by it. But water-power is only possible so long as the supply of water continues. The supply of water is dependent upon the rains; the rains depend upon evaporation; and evaporation is maintained by solar heat. The energy of *water-power* is, therefore, transformed solar energy.

A moving mass of air possesses energy equal to the mass multiplied by half the square of the velocity. This energy is available for propelling ships, for turning windmills, and for other work. Winds are due to a disturbance of atmospheric equilibrium by solar heat; and the energy of *wind-power*, like that of water-power, is, therefore, derived from solar energy.

The *ocean currents* also possess energy due to their motion, and this motion is, like that of the winds, derived from solar energy.

By far the largest part of the energy employed by man for his purposes is derived from the combustion of wood and coal. This energy exists as the potential *energy of chemical combination* of oxygen with carbon and hydrogen. Now, we know that vegetable matter is formed by the action of the solar rays through the mechanism of the leaf, and that coal is the carbon of plants that grew and decayed in a past geological age. The energy of wood and coal is, therefore, the transformed energy of solar radiations.

It is well known that, in the animal tissues, a chemical action takes place similar to that involved in combustion. The oxygen taken into the lungs and absorbed by the blood combines, by processes with which we are not here concerned, with the constituents of the food. Among the products of this combination are carbon dioxide and water, as in the combustion of the same substances elsewhere. Lavoisier assumed that such chemical combinations were the source of *animal heat*, and was the first to attempt a measurement of it. He compared the heat developed with that due to the formation of the carbon dioxide exhaled in a given

time. Despretz and Dulong made similar experiments with more perfect apparatus, and found that the heat produced by the animal was about one-tenth greater than would have been produced by the formation by combustion of the carbonic acid and water exhaled.

These and similar experiments, although not taking into account all the chemical actions taking place in the body, leave no doubt that animal heat is due to atomic and molecular changes within the body.

The work performed by muscular action is also the transformed energy of food. Rumford, in 1798, saw this clearly; and he showed, in a paper of that date, that the amount of work done by a horse is much greater than would be obtained by using its food as fuel for a steam-engine.

Mayer, in 1845, held that an animal is a heat-engine, and that every motion of the animal is a transformation into work of the heat developed in the tissues.

Hirn, in 1858, executed a series of interesting experiments bearing upon this subject. In a closed box was placed a sort of tread-mill, which a man could cause to revolve by stepping from step to step. He thus performed work which could be measured by suitable apparatus outside the box. The tread-wheel could also be made to revolve backward by a motor placed outside, when the man descended from step to step, and work was performed upon him.

Three distinct experiments were performed; and the amount of oxygen consumed by respiration, and the heat developed, were determined.

In the first experiment the man remained in repose; in the second he performed work by causing the wheel to revolve; in the third the wheel was made to revolve backward, and work was performed upon him. In the second experiment the amount of heat developed for a gram of oxygen consumed was much less, and in the third case much greater, than in the first; that is, in the first case, the heat developed was due to a chemical action, indicated by the absorption of oxygen; in the second, a portion of the chemical action went to perform the work, and hence a less amount of heat

was developed; while in the third case the motor, causing the tread-wheel to revolve, performed work, which produced heat in addition to that due to the chemical action.

It has been thought that muscular energy is due to the waste of the muscles themselves; but experiments show that the waste of nitrogenized material is far too small in amount to account for the energy developed by the animal; and we must, therefore, conclude that the principal source of muscular energy is the oxidation of the non-nitrogenized material of the blood by the oxygen absorbed in respiration.

An animal is, then, a machine for converting the potential energy of food into mechanical work: but he is not, as Mayer supposed, a heat-engine; for he performs far more work than could be performed by a perfect heat-engine, working between the same limits of temperature, and using the food as fuel.

The food of animals is of vegetable origin, and owes its energy to the solar rays. Animal heat and energy are, therefore, the transformed energy of the sun.

The *tides* are mainly caused by the attraction of the moon upon the waters of the earth. If the earth did not revolve upon its axis, or, rather, if it always presented one face to the moon, the elevated waters would remain stationary upon its surface, and furnish no source of energy. But as the earth revolves the crest of the tidal wave moves apparently in the opposite direction, meets the shores of the continents, and forces the water up the bays and rivers, where energy is wasted in friction upon the shores or may be made use of for turning mill-wheels. It is evident that all the energy derived from the tides comes from the rotation of the earth upon its axis; and a part of the energy of the earth's rotation is, therefore, being dissipated in the heat of friction it causes.

The internal heat of the earth and a few other forms of energy, such as that of native sulphur, iron, etc., are of little consequence as sources of useful energy. They may be considered as the remnants of the original energy of the earth.

236. Energy of the Sun.—It has been seen that the sun's rays

are the source of all the forms of energy practically available, except that of the tides. It has been estimated that the heat received by the earth from the sun each year would melt a layer of ice over the entire globe a hundred feet in thickness. This represents energy equal to one horse-power for each fifty square feet of surface, and the heat which reaches the earth is only one twenty-two-hundred-millionth of the heat that leaves the sun. Notwithstanding this enormous expenditure of energy, Helmholtz and Thomson have shown that the nebular hypothesis, which supposes the solar system to have originally existed as a chaotic mass of widely separated gravitating particles, presents to us an adequate source for all the energy of the system. As the particles of the system rush together by their mutual attractions, heat is generated by their collision; and after they have collected into large masses, the condensation of these masses continues to generate heat.

237. Dissipation of Energy.—It has been seen that only a fraction of the energy of heat is available for transformation into other forms of energy, and that such transformation is possible only when a difference of temperature exists. Every conversion of other forms of energy into heat puts it in a form from which it can be only partially recovered. Every transfer of heat from one body to another, or from one part to another of the same body, tends to equalize temperatures, and to diminish the proportion of energy available for transformation. Such transfers of heat are continually taking place; and, so far as our present knowledge goes, there is a tendency toward an equality of temperature, or, in other words, a uniform molecular motion, throughout the universe. If this condition of things were reached, although the total amount of energy existing in the universe would remain unchanged, the possibility of transformation would be at an end, and all activity and change would cease. This is the doctrine of the dissipation of energy to which our limited knowledge of the operations of Nature leads us; but it must be remembered that our knowledge is very limited, and that there may be in Nature the means of restoring the differences upon which all activity depends.

MAGNETISM AND ELECTRICITY.

CHAPTER I.

MAGNETISM.

238. Fundamental Facts.—Masses of iron ore are sometimes found which possess the property of attracting pieces of iron and a few other substances. Such masses are called *natural magnets* or lodestones. A bar of steel may be so treated as to acquire similar properties. It is then called a *magnet.* Such a magnetized steel bar may be used as fundamental in the investigation of the properties of magnetism.

If pieces of iron or steel be brought near a steel magnet, they are attracted by it, and unless removed by an outside force they remain permanently in contact with it. While in contact with the magnet, the pieces of iron or steel also exhibit magnetic properties. The iron almost wholly loses these properties when removed from the magnet. The steel retains them and itself becomes a magnet. The reason for this difference is not fully known. It is usually said to be due to a *coercive force* in the steel. The attractive power of the original magnet for other iron or steel remains unimpaired by the formation of new magnets.

A body which is thus magnetized or which has its magnetic condition disturbed is said to be affected by *magnetic induction.*

In an ordinary bar magnet there are two small regions, near the ends of the bar, at which the attractive powers of the magnet

are most strongly manifested. These regions are called the *poles* of the magnet. The line joining two points in these regions, the location of which will hereafter be more closely defined, is called the *magnetic axis.* An imaginary plane drawn normal to the axis at its middle point is called the *equatorial plane.*

If the magnet be balanced so as to turn freely in a horizontal plane, the axis assumes a direction which is approximately north and south. The pole toward the north is usually called the north or positive pole; that toward the south, the south or negative pole.

If two magnets be brought near together, it is found that their like poles repel and unlike poles attract one another.

If the two poles of a magnet be successively placed at the same distance from a pole of another magnet, it is found that the forces exerted are equal in amount and oppositely directed.

The direction assumed by a freely suspended magnet shows that the earth acts as a magnet, and that its north magnetic pole is situated in the southern hemisphere.

If a bar magnet be broken, it is found that two new poles are formed, one on each side of the fracture, so that the two portions are each perfect magnets. This process of making new magnets by subdivision of the original one may be, so far as known, continued until the magnet is divided into its least parts, each of which will be a perfect magnet.

This last experiment enables us at once to adopt the view that the properties of a magnet are due to the resultant action of its constituent magnetic molecules.

239. Law of Magnetic Force.—By the help of the torsion balance, the principle of which is described in §§ 109, 253, and by using very long, thin, and uniformly magnetized bars, in which the poles can be considered as situated at the extremities, Coulomb showed that the repulsion between two similar poles, and the attraction between two dissimilar poles, is inversely as the square of the distance between them.

A more exact proof of the same law was given by Gauss, who calculated the action of one magnet on another on the assumption

of the truth of the law, and showed by experiment that the action calculated was actually exerted.

All theories of magnetism assume that the force between two magnet poles is proportional to the product of the strengths of the poles. The law of magnetic force is then the same as that upon which the discussion of potential and of flux of force was based. The theorems there discussed are in general applicable in the study of magnetism, although modifications in the details of their application occur, arising from the fact that the field of force about a magnet is due to the combined action of two dissimilar and equal poles.

If m and m' represent the strengths of two magnet poles, r the distance between them, and k a factor depending on the units in which the strength of the pole is measured, the formula expressing the force between the poles is $k\frac{mm'}{r^2}$.

240. Definitions of Magnetic Quantities.—The law of magnetic force enables us to define a *unit magnet pole*, based upon the fundamental mechanical units.

If two perfectly similar magnets, infinitely thin, uniformly and longitudinally magnetized, be so placed that their positive poles are unit distance apart, and if these poles repel one another with unit force, the magnet poles are said to be of *unit strength.* Hence, in the expression for the force between two poles, k becomes unity, and the dimensions of $\frac{m^2}{r^2}$ are those of a force. That is, $\left[\frac{m^2}{r^2}\right] = MLT^{-2}$, from which the dimensions of a magnet pole are $[m] = M^{\frac{1}{2}}L^{\frac{3}{2}}T^{-1}$. This definition of a unit magnet pole is the foundation of the *magnetic system of units.* The *strength* of a magnet pole is then equal to the force which it will exert on a unit pole at unit distance.

The product of the strength of the positive pole of a uniformly and longitudinally magnetized magnet into the distance between its poles is called its *magnetic moment.*

The quotient of the magnetic moment of such a magnet by its volume, or the magnetic moment of unit of volume, is called the *intensity of magnetization.* Since any magnet may be divided into small magnets, each of which is uniformly magnetized, and for which by this definition a particular value of the intensity of magnetization can be found, it is clear that the magnetic condition of any magnet can be stated in terms of the intensity of magnetization of its parts.

The dimensions of magnetic moment and of intensity of magnetization follow from these definitions. They are respectively

$$[ml] = M^{\frac{1}{2}}L^{\frac{5}{2}}T^{-1} \quad \text{and} \quad \left[\frac{ml}{l^3}\right] = M^{\frac{1}{2}}L^{-\frac{1}{2}}\,T^{-1}.$$

241. Distribution of Magnetism in a Magnet.—If we conceive of a single row of magnetic molecules with their unlike poles in contact, we can easily see that all the poles, except those at the ends, neutralize one another's action, and that such a row will have a free north pole at one end and a free south pole at the other. If a magnet be thought of as made up of a combination of such rows of different lengths, the action of their free poles may be represented by supposing it due to a distribution of equal quantities of two imaginary substances, called north and south *magnetism.* This distribution will be, in general, both on the surface and throughout the volume of the magnet. If the magnet be uniformly magnetized, the volume distribution becomes zero. The *surface distribution* of magnetism will sometimes be used to express the magnetization of a magnet, by the use of a concept called the *magnetic density.* It is defined as the ratio of the quantity of magnetism on an element of surface to the area of that element. The magnetic density thus defined has the same numerical value as the intensity of magnetization which measures the real distribution. To illustrate this statement, we will consider an infinitely thin and uniformly magnetized bar, of which the length and cross-section are represented by l and s respectively. Its intensity of magnetization is $\frac{ml}{ls}$ or $\frac{m}{s}$. If, now, for the pole m we sub-

stitute a continuous surface distribution over the end of the bar, then $\frac{m}{s}$ is also the density of that distribution.

The dimensions of magnetic density follow from this definition. They are $\left[\frac{m}{s}\right] = \frac{M^{\frac{1}{2}}L^{\frac{3}{2}}T^{-1}}{L^2} = M^{\frac{1}{2}}L^{-\frac{1}{2}}T^{-1}$.

Coulomb showed, by oscillating a small magnet near different parts of a long bar magnet, that the magnetic force at different points along it gradually increases from the middle of the bar, where it is imperceptible, to the extremities. This would not be the case if the bar magnet were made up of equal straight rows of magnetic molecules in contact, placed side by side. With such an arrangement there would be no force at any point along the bar, but it would all appear at the two ends. The mutual interaction of the molecules of contiguous rows makes such an arrangement, however, impossible.

In the earth's magnetic field, in which the lines of magnetic force may be considered parallel, a couple will be set up on any magnet, so magnetized as to have only two poles, due to the action of equal quantities of north and south magnetism distributed in the magnet. The points at which the forces making up this couple are applied are the *poles* of the magnet, and the line joining them is the *magnetic axis.* These definitions are more precise than those which could be given at the outset.

242. Action of One Magnet on the Other.—The investigation of the mechanical action of one magnet on another is important in the construction of apparatus for the measurement of magnetism.

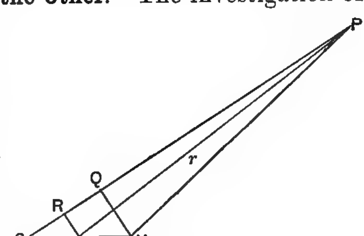

FIG. 74.

(1) To determine the *potential of a short bar magnet* at a point distant from it, let NS (Fig. 74) represent the magnet of length $2l$, the poles of which are of strength m, and let the point P be at a distance r from the centre of the magnet, taken as origin.

Let the angle PON equal θ and draw the perpendiculars NQ and OR to PS. Then, in the limit, if SN is very small in comparison with OP, we have $PN = r - \Delta r$ and $PS = r + \Delta r$, where Δr is a small length equal to $SR = l \,.\, \cos\theta$. The potential at P due to the pole at N is $\dfrac{m}{r - \Delta r} = m\left(\dfrac{1}{r} + \dfrac{\Delta r}{r^2}\right)$, since Δr is very small in comparison with r. Similarly the potential at P due to the pole at S is $-\dfrac{m}{r + \Delta r} = -m\left(\dfrac{1}{r} - \dfrac{\Delta r}{r^2}\right)$. The potential at P due to the magnet is therefore

$$\frac{2m\Delta r}{r^2} = \frac{2ml\cos\theta}{r^2} = \frac{M\cos\theta}{r^2}, \tag{81}$$

where M is the magnetic moment of the magnet. We may consider the magnetic moment as projected upon the line r by multiplication by $\cos\theta$; the formula shows that the potential at any point due to a short magnet is equal to the projection of the magnetic moment upon the line joining the centre of the magnet with the point, divided by the square of the length of that line.

The maximum value of the potential due to the magnet, for a given value of r, is $\dfrac{M}{R^2}$, where R represents the assigned value of r. If we set $\dfrac{M}{R^2} = \dfrac{M \,.\, \cos\theta}{r^2}$ we obtain $r^2 = R^2\cos\theta$ as the equation of the equipotential surfaces at a considerable distance from the small magnet. When $R = \infty$, it determines an equipotential surface of zero potential, for which, for every finite value of r, we have $\cos\theta = 0$, and $\theta = \dfrac{\pi}{2}$. The plane passing through the centre of the magnet and perpendicular to its axis is therefore an equipotential surface of zero potential. Since $r = 0$ whenever $\cos\theta = 0$, whatever be the value of R, all the other equipotential surfaces pass through the point O; they are in general ovoid surfaces surrounding the poles. The lines of force of the magnet arise at the north pole and pass perpendicularly through all these surfaces to the south pole.

(2) *The force due to a short bar magnet* in any direction may be determined by determining the rate of change of its potential in that direction. It is not, however, important to determine this force in the general case: it will be sufficient to determine it for points in the line of the axis of the magnet.

Let the length of the magnet NS (Fig. 75) be represented by $2l$ and the distance from its centre O to the point P by r. Then the force at P due to the pole at N, and directed away from the magnet, is $\frac{m}{(r-l)^2}$, and the force due to the pole at S, and directed toward the magnet, is $\frac{m}{(r+l)^2}$. Now we may write

FIG. 75.

$\frac{m}{(r-l)^2} = \frac{m}{r^2 - 2lr} = m\left(\frac{1}{r^2} + \frac{2l}{r^3}\right)$, since l is very small in comparison with r, and similarly $\frac{m}{(r+l)^2} = m\left(\frac{1}{r^2} - \frac{2l}{r^3}\right)$. The force at P due to the magnet and directed away from it is, therefore,

$$\frac{4ml}{r^3} = \frac{2M}{r^3}. \tag{82}$$

(3) In the construction of apparatus used in the measuring of magnetic quantities it is important to know the *moment of couple* set up by one magnet on another. We will determine this for the particular case in which both the magnets are small in comparison with the distance between their centres, and in which the centre of one is situated on the prolongation of the axis of the other. We will call the magnet, the axis of which lies in the line joining the centres, the first magnet, and the other the second magnet, and will examine the couple exerted on the second magnet by the first. Under the limitations made as to the size of the magnets, we may assume that the forces exerted by the first magnet on the poles of the second are the same as if the poles of the second magnet lay in the prolongation of the axis of the first magnet, and that they are the same for any position of the second magnet (Fig. 76).

We designate by m' the pole of the second magnet, by $2l'$ its length, and by θ the complement of the angle made by its axis with the line joining the centres of the magnets. On these assumptions, the force acting on the north pole of the second magnet is $\frac{2m'M}{r^3}$, and the force acting on its south pole is $-\frac{2m'M}{r^3}$. These two forces constitute a couple with an arm $2l' \cos\theta$, and the moment of this couple is

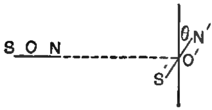

FIG. 76.

$$\frac{4m'l'M \cos\theta}{r^3} = \frac{2MM' \cos\theta}{r^3}, \tag{83}$$

where M' represents the magnetic moment of the second magnet. This moment of couple varies from $\frac{2MM'}{r^3}$ if the magnets are at right angles to each other, to zero if they are in the same straight line.

243. The Magnetic Shell.—A *magnetic shell* may be defined as an infinitely thin sheet of magnetizable matter, magnetized transversely; so that any line in the shell normal to its surfaces may be looked on as an infinitesimally short and thin magnet. These imaginary magnets have their like poles contiguous. The product of the intensity of magnetization at any point in the shell into the thickness of the shell at that point is called the *strength of the shell* at that point, and is denoted by the symbol j.

Since we may substitute for the magnetic arrangement an imaginary distribution of magnetism over the surfaces of the shell, we may define the strength of the shell as the product of the surface-density and the thickness of the shell.

The dimensions of the strength of a magnetic shell follow at once from this definition. We have $[j]$ equal to the dimensions of intensity of magnetization multiplied by a length. Therefore $[j] = M^{\frac{1}{2}}L^{\frac{1}{2}}T^{-1}$.

We obtain first the potential of such a shell of infinitesimal

area. Let the origin (Fig. 77) be taken half-way between the two faces of the shell, and let the shell stand perpendicular to the x axis. Let a represent the area of the shell, supposed infinitesimal, $2l$ the thickness of the shell, and d the intensity of magnetization. The volume of this infinitesimal magnet is $2al$, and, from the definition of intensity of magnetization, $2ald$ is its magnetic moment. The potential at the point P is then given by equation (81), since l is very small. We have $V = \frac{M}{r^2}\cos\theta = \frac{2ald}{r^2}\cos\theta$.

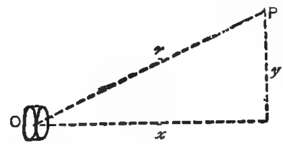

FIG. 77.

Now $a\cos\theta$ is the projection of the area of the shell upon a plane through the origin normal to the radius vector r, and, since a is infinitesimal, $\frac{a\cos\theta}{r^2}$ is the solid angle ω bounded by the lines drawn from P to the boundary of the area a. The potential then becomes $V = 2ld\omega = j\omega$, since $2ld$ is what has been called the strength of the shell.

The same proof may be extended to any number of contiguous areas making up a finite magnetic shell. The potential due to such a shell is then $\Sigma j\omega$. If the shell be of uniform strength, the potential due to it becomes $j\Sigma\omega$, and is got by summing the elementary solid angles. This sum is the solid angle Ω, bounded by the lines drawn from the point of which the potential is required to the boundary of the shell. The potential due to a magnetic shell of uniform strength is therefore

$$j\Omega. \tag{84}$$

It does not depend on the form of the shell, but only on the angle subtended by its contour. At a point very near the positive face of a flat shell, so near that the solid angle subtended by the shell equals 2π, the potential is $2\pi j$; at a point in the plane of the shell outside its boundary, where the angle subtended is zero, the potential is zero; and near the other or negative face of the shell it is $-2\pi j$. The whole work done, then, in moving a unit magnet pole from a point very near one face to a point very near the other

face is $4\pi j$. This result is of importance in connection with electrical currents.

244. Magnetic Measurements.—It was shown by Gilbert in a work published in 1600, that the earth can be considered as a magnet, having its positive pole toward the south and its negative toward the north. The determination of the magnetic relations of the earth are of importance in navigation and geodesy. The principal magnetic elements are the declination, the dip, and the horizontal intensity.

The *declination* is the angle between the magnetic meridian, or the direction assumed by the axis of a magnetic needle suspended to move freely in a horizontal plane, and the geographical meridian.

The *dip* is the angle made with the horizontal by the axis of a magnetic needle suspended so as to turn freely in a vertical plane containing the magnetic meridian.

The *horizontal intensity* is the strength of the earth's magnetic field resolved along the horizontal line in the plane of the magnetic meridian. A magnet pole of strength m in a field in which the horizontal intensity is represented by H is urged along this horizontal line with a force equal to mH. From this equation the dimensions of the horizontal intensity, and so also of the strength of a magnetic field in any case, are $[H] = \left[\frac{MLT^{-2}}{m}\right] = M^{\frac{1}{2}}L^{-\frac{1}{2}}T^{-1}$.

The horizontal intensity can be measured relatively to some assumed magnet as standard, by allowing the magnet to oscillate freely in the horizontal plane about its centre, and noting the time of oscillation. The relation between the magnetic moment M of the magnet and the horizontal intensity H is calculated by a formula analogous to that employed in the computation of g from observations with the pendulum.

If the magnet be slightly displaced from its position of equilibrium, so as to make small oscillations about its point of suspension, it can be shown, as in § 60, that it is describing a simple harmonic motion. If ϕ represent the angle made by the magnet with the magnetic meridian, the moment of couple acting on the magnet is

given by $MH \sin \phi = MH\phi$, if the oscillations are always very small. If I represent the moment of inertia of the magnet, we have (§ 39) $MH\phi = I\alpha$, where α is the angular acceleration. Now since the motion of each particle of the magnet is simple harmonic, and since the linear motions of the particles are proportional to their angular motions, we have $\alpha = \frac{4\pi^2}{T^2}\phi$, and by substituting this value of α, we obtain

$$MH = \frac{4\pi^2}{T^2} I. \tag{85}$$

The moment of inertia I may be either computed directly from the magnet itself, if it be of symmetrical form, or it may be determined experimentally by the method given in § 37, which applies in this case. The horizontal intensity is then determined relative to the magnetic moment of the assumed standard magnet.

This result may be used to give an absolute measure of H by combining with it the result of another observation which gives an independent relation between M and H. In the arrangement of the apparatus two magnets are used: one, the *deflected magnet*, is so suspended as to turn freely in the horizontal plane; and the other, the *deflecting magnet*, the one of moment M used in the last operation, is carried upon a bar which can be set at right angles to the magnetic meridian. The centre of the deflected magnet is in the prolongation of the axis of the deflecting magnet. The deflected magnet makes an angle θ with the magnetic meridian determined by the equality between the two couples acting on the deflected magnet, one arising from the action of the earth's magnetism, and the other from that of the deflecting magnet. This latter has been already discussed in § 242.

The couple due to the deflecting magnet is given by $\frac{2MM'}{r^3} \cos \theta$, and that due to the earth's magnetism by $M'H \sin \theta$. We have then

$$\frac{2MM'}{r^3} \cos \theta = M'H \sin \theta, \quad \text{or} \quad \frac{M}{H} = \tfrac{1}{2} r^3 \tan \theta. \tag{86}$$

Equations (85) and (86) contain the unknown quantities M and H

in different relations. By the elimination of one of them the other can be obtained in absolute units. In practice the simple conditions assumed in this discussion cannot be obtained, and corrections must be introduced, arising from the departures from these conditions. In the determination of MH we must take into account the change of the magnetic moment by the induction of the field, and the facts that the oscillations are not infinitesimal, and that they are affected by the friction of the air and the torsion of the suspension fibre. In the determination of $\frac{M}{H}$, we must take into account the induction of one magnet on the other, and the fact that the lengths of the magnets are not negligible in comparison with the distance between them.

245. The Magnetic Field.—Up to this point our discussion has been conducted on the supposition that forces obeying a definite law act directly between magnetic poles. Various phenomena, especially those of magnetic induction and the relation between magnetism and the electrical current, as well as the general tendency of modern speculation in physics, lead us to think that this mode of representing the interactions of magnets is an artificial one, and that the true seat of the magnetic action is in the medium between the magnets. From this point of view it is desirable to have a mode of describing the magnetic field in such a way that the relation between its condition and the magnetic forces exhibited may be expressed in measurable terms. The most useful mode of describing the field is that depending upon the use of lines and tubes of force.

Since the force due to a magnet pole obeys the law of inverse squares, the theorems of §§ 53–57 are immediately applicable.

For the sake of clearness in statement we will define a *unit tube* of force in the following way: Consider a closed surface which encloses a quantity of free positive magnetism m. By § 56, $\Sigma Fs = 4\pi m$, where F represents the normal component of the force at each point on the surface and s the area of the element of surface over which it acts. Now let us suppose the whole surface divided into $4\pi m$ parts, for each of which $Fs = 1$; then the tubes

of force determined by the contours of these areas will be called unit tubes. Since $\frac{1}{s} = N$ is the number of unit tubes which pass through a unit area, we have $N = F$. Hence the magnetic force at a point is equal to the number of unit tubes which pass perpendicularly through unit area at that point.

246. Magnetic Force and Magnetic Induction.—If a body like a piece of iron be brought into a magnetic field it will be magnetized by induction, and will in turn affect the distribution of the tubes of force in the field. To determine the way in which its magnetization is affected, we need some convention upon which we shall measure the force within it. This force is measured within a cavity in the mass of iron, and will have different values for different forms of the cavity. Let us first suppose the cavity cylindrical, with its axis in the direction of the lines of force, and with its bases infinitesimal in comparison with its height. The distribution of magnetism throughout the iron will give rise to free magnetism on the bases of the cylinder, and the force on a unit pole placed at the middle of the cylinder will be due to the original force in the field, to the forces arising from the induced poles, and to the forces exerted by these distributions. These last forces are infinitesimal, in case the bases are infinitesimal in comparison with the length of the cylinder, and may be neglected. The force on the pole is called the *magnetic force* at the point where the pole is placed.

Let us in the second place suppose the cavity in the shape of a disk, with its faces normal to the lines of force, and infinitely great in comparison with the distance between them. The magnetization of the iron will give rise to a distribution of free magnetism on each of these faces. If we consider the lines of force of the field as passing into the iron from left to right, the distribution on the left-hand face of the cavity is positive and that on the right-hand face negative. The force on a unit pole placed within the cavity is due to the original force R in the field, and the forces exerted by the distributions on the faces of the cavity. If we represent by σ the density of this distribution on either face, the force on the pole

due to each face is $2\pi\sigma$ (§ 57), and the forces from the two faces act in the same direction. The force upon the pole due to the faces of the cavity is therefore $4\pi\sigma$ or $4\pi I$, where I is the intensity of magnetization (§ 241). The total force acting on the pole is therefore

$$F = R + 4\pi I; \tag{87}$$

this quantity is called the *magnetic induction* within the body. It is manifestly a directed quantity or a vector; in the example considered its direction is the same as that of the force. In bodies which are not isotropic, the induction and the magnetizing force are not necessarily in the same direction. We will confine our attention to isotropic bodies.

247. Tubes of Force of a Magnet. Tubes of Induction.—The lines of force in the field outside a bar magnet are curves proceeding from the north to the south pole; these lines of force may be conceived of as existing also within the body of the magnet if the magnetic force within the magnet is determined within the long narrow cylinder already described. On the other hand, a set of tubes may be described, called *tubes of induction*, which are closed tubes, proceeding through the magnet from the south to the north pole and outside the magnet from the north to the south pole. To show this let us suppose the magnet divided into two parts by a section at right angles to its axis, and let us consider a closed surface passing between the two parts and enclosing that part which contains the north pole. The distribution over the end of this part which is exposed by the section is equal to the north pole at the end of the original magnet, and is of opposite sign; so that the flux of force $\Sigma Fs = 0$ over the closed surface. Now the force within the cavity formed by the section is directed inward, and is at each point equal to $4\pi I$. If a represent the area of the section, the flux of force across that part of the surface contained within the section is $4\pi Ia$, and $Ia = m$, the strength of the pole of the original magnet. The number of tubes of force which pass through the section of the magnet is therefore equal to $4\pi m$, or to the number of tubes of force which proceed from the original pole and pass

through the rest of the closed surface from within outward. The tubes of force of the magnet are therefore closed tubes, passing through the magnet in the manner already described. The tubes thus considered, in which account is taken of the effect of the distributions on the disk-shaped cavity within the magnet, are called *tubes of induction.*

If the magnet be a bar placed with its axis along the lines of force in a uniform magnetic field and magnetized by induction, the *induction* (§ 246) is equal to $F = R + 4\pi I$. If the magnetization of the bar be proportional to the force of the field, so that $I = kR$, we have

$$F = (1 + 4\pi k)R = \mu R. \tag{88}$$

The number of tubes of induction which pass through unit area in a cross-section of the bar is equal to this, for the total number of tubes that pass through the section of the magnet is $(R + 4\pi I)a$; that is, $N = \mu R$.

The coefficient k is called the *coefficient of induced magnetization;* it is assumed to be zero for a vacuum, and may be either positive or negative. The coefficient $\mu = 1 + 4\pi k$ is called the *magnetic inductive capacity* or the *magnetic permeability.* It must be noticed that when $k > 0$, so that the induction is greater than the magnetic force of the field, the resultant magnetic force within the body is less than the magnetic force of the field, because the poles induced in the body act in the opposite sense to the force of the field.

248. Energy in a Magnetic Field.—On the view we are now taking, that the actions between magnets are due to a condition of the medium which occupies the field, it is natural to suppose that the energy of a set of magnets is distributed in the field. We will find a law for this distribution, which associates the energy with the tubes of induction.

The energy of the system is manifestly equal to the work that would be required to construct that system. We will first show that this may be expressed, in terms of the magnet poles and of the potentials of the places occupied by them, by the formula $\Sigma\frac{1}{2}mV$.

We assume that whatever bodies are in the field are of such a character that their magnetization is proportional to the magnetizing force; on this assumption, the potential at any point and the magnitude of the poles vary in the same proportion. Let m_1, m_2, $\ldots . m_n$ represent the values of the respective poles, and V_1, V_2, $\ldots . V_n$ the potentials at the places occupied by them in the final condition of the field. Each of these poles may be conceived of as an assemblage of a great number n of small poles, each equal to $\frac{m}{n}$. If we think of the region occupied by the field as originally free from magnets, its energy after the magnets are present in it will be equal to the work done in forming the magnet poles by the successive addition of such elementary poles. Let the field be free from magnetism, and let the quantities of magnetism $\frac{m_1}{n}$, $\frac{m_2}{n}$, $\ldots . \frac{m_n}{n}$, be brought to the points which the separate poles occupy in the final condition of the field; since the potentials at those points are originally zero, no work will be done in this operation. The presence of these poles causes a rise of potential throughout the field, and the potentials at the places occupied by the poles become $\frac{V_1}{n}$, $\frac{V_2}{n}$, $\ldots . \frac{V_n}{n}$. Let elementary poles similar to those already introduced be brought to their respective places in the field; the work done on any one of them is $\frac{mV}{n^2}$, and the work done on them all is $\Sigma\frac{mV}{n^2}$. By this increase in the quantities at the poles the potentials become $2\frac{V_1}{n}$, $2\frac{V_2}{n}$, $\ldots . 2\frac{V_n}{n}$. This operation is repeated until n quantities have been brought to each pole, so that the poles are in their final condition and the potential has everywhere its final value. The work done in bringing up the n^{th} elementary pole to its place is $\frac{m}{n}(n-1)\frac{V}{n}$; the work done in forming the field is therefore $\Sigma\left(\frac{1+2+3+\ldots+(n-1)}{n^2}\right)mV$. Now

$$\frac{1+2+3+\ldots+(n-1)}{n^2} = \frac{(n-1)n}{2n^2} = \frac{1}{2}\left(1-\frac{1}{n}\right) = \frac{1}{2},$$

if n be supposed to be very large. The work done in forming the magnetic field is therefore $\Sigma\frac{1}{2}mV$.

Now, to show how this energy may be distributed in the field, we may consider any one of the magnets which give rise to the field as being the origin of $4\pi Ia$ unit tubes of induction, the magnets being thought of as bar magnets. The energy of this magnet is, by the previous proposition, equal to $\frac{1}{2}m(V_n - V_s)$, where V_n and V_s are the potentials at the places occupied by its poles; the pole m is equal to Ia (§ 241). The difference of potential $V_n - V_s$ equals $\Sigma R\Delta l$, where R is the force along a line of force in the field passing outside the magnet from its north to its south pole, and Δl is an element of that line, the summation being extended over the whole line. If, therefore, we suppose each unit tube of induction which proceeds from the magnet to contain an amount of energy equal to $\Sigma\frac{R\Delta l}{8\pi}$, the energy contained in the bundle of tubes belonging to the magnet will equal the energy of the magnet. We may therefore consider the energy of the magnet as distributed throughout the field, in such a way that each unit length of a unit tube of induction contains $\frac{R}{8\pi}$ units of energy. The tubes of induction here considered are those which exist outside the magnets. It has already been shown that the number of tubes of induction which pass through unit area is equal to the induction, or that $N = F = (1 + 4\pi k)R = \mu R$. Hence the energy in unit length of a tube of induction may be expressed by $\frac{N}{\mu 8\pi}$.

The energy in unit volume of the field may be determined by considering a small cylinder of length l and cross-section s placed in the field with its end surfaces normal to the lines of induction. The number of tubes of induction which pass through the end surfaces is $Ns = \mu Rs$, and the energy contained in the length l of each of these tubes is $\frac{Nl}{\mu 8\pi} = \frac{Rl}{8\pi}$. The energy contained

in the cylinder is therefore $\frac{N^2ls}{\mu 8\pi} = \frac{\mu R^2 ls}{8\pi}$, and the energy contained in unit volume is $\frac{N^2}{\mu 8\pi} = \frac{\mu R^2}{8\pi}$.

249. Paramagnetism and Diamagnetism.—It was discovered by Faraday that all bodies are affected when brought into a magnetic field: some of them, such as iron, nickel, cobalt, and oxygen, are attracted by the magnet setting up the field; others, such as bismuth, copper, most organic substances, and nitrogen, are repelled from the magnet. The former are said to be *ferromagnetic* or *paramagnetic*, the latter *diamagnetic.*

The most obvious explanation of these phenomena, and the one adopted by Faraday, is to ascribe them to a distribution of the induced magnetization in paramagnetic bodies, in an opposite direction from that in diamagnetic bodies. If a paramagnetic body be brought between two opposite magnet poles, a north pole is induced in it near the external south pole, and a south pole near the external north pole. The magnetic separation is then said to be in the direction of the lines of force. According to this explanation, then, the separation of the induced magnetization in a diamagnetic body is in a direction opposite to that of the lines of force. In other words, if a diamagnetic body be brought between two opposite magnet poles, the explanation asserts that a north pole is induced in it near the external north pole, and a south pole near the external south pole.

One of Faraday's experiments, however, indicates that the different behavior of bodies of these two classes may be due only to a more or less intense manifestation of the same action. He found that a solution of ferrous sulphate, sealed in a glass tube, behaves, immersed in a weaker solution of the same salt, as a paramagnetic body; but, when immersed in a stronger solution, as a diamagnetic body. It may from this experiment be concluded that the direction of the induced magnetization is the same for all bodies, and that the exhibition of diamagnetic or paramagnetic properties depends, not upon the direction of induced magnetization, but upon

the greater or less intensity of magnetization of the surrounding medium.

Faraday discovered that many bodies while in a vacuum exhibit diamagnetic properties. In accordance with this explanation, we must conclude that a vacuum can have magnetic properties. It seemed to Faraday unlikely that this should be the case, and he therefore adopted the explanation which was first given. As it has since been shown that the ether which serves as a medium for the transmission of light, and which pervades every so-called vacuum, is also probably concerned in electrical and magnetic phenomena, there is no longer any reason for the opinion that the possession of magnetic properties by a vacuum is inherently improbable.

To classify bodies as paramagnetic or diamagnetic, we examine the energy existing in them when placed in a magnetic field. We will first assume that k, the coefficient of magnetization, is so small that the resultant force in the region occupied by the body is not appreciably changed by the presence of the body. The value of k for vacuum is assumed to be zero, and for air it is very slightly different from zero; hence the value of μ for air may be set equal to 1. Before a body is brought into the field, the energy per unit volume in the space finally occupied by it is $\frac{N^2}{8\pi}$; the energy per unit volume in the same space when the body is brought into the field is $\frac{N^2}{\mu 8\pi}$. The increase of energy caused by the introduction of the body, on the assumption we have made that the field is not disturbed by the body, or that N remains the same after the introduction of the body as it was before, is $\frac{N^2}{8\pi}\left(\frac{1-\mu}{\mu}\right)$, and this is positive or negative according as μ is less or greater than 1. Now a body free to move will move so as to diminish its potential energy, and therefore a body for which $\mu > 1$ will move so as to make N^2 as large as possible, or will move from a weaker to a stronger part of the field. Such a body is called a paramagnetic body. On the other hand, a body for which $\mu < 1$ will move so

as to make N^2 as small as possible, that is, will move from a stronger to a weaker part of the field. Such a body is a diamagnetic body. These motions are not necessarily along the lines of force, but are in the directions in which N^2 changes most rapidly.

In general, the introduction of a body into a field changes the arrangement of the tubes of induction and the force in the field; it has already been remarked that the resultant force at a point is diminished by the presence of a paramagnetic body around that point, because the induced distribution acts against the forces in the field. Since the energy per unit length of the tubes of induction is equal to $\frac{R}{8\pi}$, where R is the resultant force, a tube of induction within a paramagnetic body possesses less energy per unit length than it does outside that body. In accordance with the tendency of the potential energy to become a minimum, the tubes of induction will therefore move into a paramagnetic body. On the other hand, the resultant force at a point being increased by the presence of a diamagnetic body around that point, the tubes of induction will move out from a diamagnetic body. This movement of the tubes of induction, into or out of bodies in the field, ceases when the loss of potential energy due to their movement into or out of the body is balanced by the gain in potential energy due to the lengthening of those tubes of the field which do not pass through the body. This change in the arrangement of the tubes of induction does not invalidate the former conclusion that paramagnetic bodies move from a place of weaker to a place of stronger magnetic force, while diamagnetic bodies move from a place of stronger to a place of weaker magnetic force. A complete discussion of the behavior of bodies in a magnetic field is outside the scope of this work.

250. Changes in Magnetization. Hysteresis.—When a magnetizable body is placed in a powerful magnetic field, it often receives, temporarily, a more intense magnetization than it can retain when removed. It is said to be *saturated*, or magnetized to *saturation*, when the intensity of its magnetization has its highest

possible value. The coercive force of steel is much greater than that of any other substance; the intensity of magnetization which it can retain is, therefore, relatively very great, and it is hence used for permanent magnets. The coercive force depends upon the quality and temper of the steel.

It was found by Ewing that the intensity of magnetization of a mass of iron in a magnetic field of given intensity is not dependent upon that intensity alone, but depends also upon the previous history of the magnetic body. In general the intensity of magnetization lags behind the magnetizing force; that is, if the magnetizing force be increasing, the intensity of magnetization is less for a given value of the force than it is for the same value if the force be diminishing. The relations between these two quantities are exhibited in Fig. 78, in which the magnetizing force is measured along the axis H, the intensity of magnetization along the axis I. The curve $ACBD$ represents the relation of these quantities as the magnetizing force of the field changes from a high negative to a high positive value. The area between these curves may be shown to measure the work done on the magnet during the cycle $ACBD$; this work is almost wholly expended in heating the magnet. The phenomenon here described is called *magnetic hysteresis*.

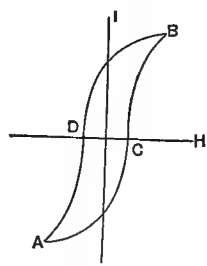

FIG. 78.

Changes of temperature cause corresponding changes in the magnetization of a magnet. If the temperature of a magnet be gradually raised, its magnetization diminishes by an amount which, for small temperature changes, is nearly proportional to the change of temperature. The magnet recovers its original magnetization when cooled again to the initial temperature, provided that the temperature to which it was raised was never very high. If it be raised, however, to a red heat, all traces of its original magnetization permanently disappear. Trowbridge has shown that, if the temperature of a magnet be carried below the tempera-

ture at which it was originally magnetized, its magnetization also temporarily diminishes.

Any mechanical disturbance, such as jarring or friction, which increases the freedom of motion among the molecules of a magnet, in general brings about a diminution of its magnetization. On the other hand, similar mechanical disturbances facilitate the acquisition of magnetism by any magnetizable body placed in a magnetic field.

251. Theories of Magnetism.—It has been shown by mathematical analysis that the facts of magnetic interactions and distribution are consistent with the hypothesis, which we have already made, that the ultimate molecules of iron are themselves magnets, having north and south poles which attract and repel similar poles in accordance with the law of magnetic force. Poisson's theory, upon which most of the earlier mathematical work is based, is that there exist in each molecule indefinite quantities of north and south magnetic fluids, which are separated and moved to opposite ends of the molecule by the action of an external magnetizing force. Weber's view, which is consistent with other facts that Poisson's theory fails to explain, is that each molecule is a magnet, with permanent poles of constant strength; that the molecules of an iron bar are, in general, arranged so as to neutralize one another's magnetic action, but that, under the influence of an external magnetizing force, they are arranged so that their magnetic axes lie more or less in some one direction. The bar is then magnetized. On this hypothesis there should be a limit to the possible intensity of magnetization, which would be reached when the axes of all the molecules have the same direction. Direct experiments by Joule, J. Müller, and Ewing indicate the existence of such a limit. An experiment of Beetz, in which a thin filament of iron deposited electrolytically in a strong magnetic field becomes a magnet of very great intensity, points in the same direction. The coercive force is, on this hypothesis, the resistance to motion experienced by the molecules. The facts that magnetization is facilitated by a jarring of the steel brought into

the magnetic field, that a bar of iron or steel after being removed from the magnetic field retains some of its magnetic properties, that the dimensions of an iron bar are altered by magnetization, the bar becoming longer and diminishing in cross-section, and that a magnetized steel bar loses its magnetism if it be highly heated, are all best explained by Weber's hypothesis.

The phenomena of hysteresis led Ewing to form an extension of Weber's theory. Ewing called attention to the fact that the neutral state of a mass of iron or other magnetizable matter cannot be due to an indiscriminate or unordered arrangement of its constituent magnetic molecules, as was assumed in Weber's form of the theory, but that the interactions of the molecular magnets occasion the formation of groups of molecules, definitely ordered, and such that they separately produce no external magnetic effect. Such groups may be roughly represented by an assemblage of small magnets floated on water or mounted on needle points, and left free to arrange themselves under the influence of their magnetic forces. If a mass of iron be placed in a field in which the magnetizing force continually increases, the first effect will be a slight development of the magnetism of the iron, due to slight changes in the directions of its molecules. When these changes have progressed so far that the original groups of molecules break down, a very rapid increase of magnetism results and new groups are formed, which are in equilibrium under the magnetic forces of the molecules and the forces of the field. As the force in the field still further increases, the increase in the magnetization of the iron still goes on, but more slowly, this increase being due to slight changes in the positions of the molecules in the new groups, which do not, however, destroy those groups. Saturation is reached, on this theory, when the molecules are arranged in parallel lines; no increase in the intensity of magnetization will then be produced by any further increase in the magnetizing force. This limit has been practically reached by Ewing in some of his experiments. If the magnetizing force be now diminished, the intensity of magnetization also diminishes, but at first only slightly. The magnetic

forces between the molecules maintain equilibrium in the molecular groups after the magnetizing force has fallen below the value at which they were formed; sufficient reduction of the magnetizing force at last occasions a breaking down of these groups and permits the formation of new ones, which exert less external magnetic force. During the time in which this change in grouping occurs, the intensity of magnetization diminishes rapidly; it does not, however, vanish until the magnetizing force has attained a finite negative value. From this point on, changes similar to those already described go on in the reverse sense. Ewing's theory also explains very well all the facts explained by Weber's form of the theory.

CHAPTER II.

ELECTRICITY IN EQUILIBRIUM.

252. Fundamental Facts.—(1) If a piece of glass and a piece of resin be brought in contact, or preferably rubbed together, it is found that, after separation, the two bodies are attracted towards each other. If a second piece of glass and a second piece of resin be treated in like manner, it is found that the two pieces of glass repel each other and the two pieces of resin repel each other, while either piece of glass attracts either piece of resin. These bodies are said to be *electrified* or *charged.*

All bodies may be electrified, and in other ways than by contact. It is sufficient for the present to consider the single example presented. The experiment shows that bodies may be in two distinct and dissimilar states of electrification. The glass treated as has been described is said to be vitreously or *positively electrified,* and the resin resinously or *negatively electrified.* The experiment shows also that bodies similarly electrified repel one another, and bodies dissimilarly electrified attract one another.

(2) If a metallic body, supported on a glass rod, be touched by the rubbed portion of an electrified piece of glass, it will become positively electrified. If it be then joined to another similar body by means of a metallic wire, the second body is at once electrified. If the connection be made by means of a damp linen thread, the second body becomes electrified, but not so rapidly as before. If the connection be made by means of a dry white silk thread, the second body shows no signs of electrification, even after the lapse of a considerable time. Bodies are divided according as they can

be classed with the metals, damp linen, or silk, as *good conductors*, *poor conductors*, and *insulators*. The distinction is one of degree. All conductors offer some opposition to the transfer of electrification, and no body is a perfect insulator under all conditions.

A conductor separated from all other conductors by insulators is said to be *insulated*. A conductor in conducting contact with the earth is said to be *grounded* or joined to ground.

During the transfer of electrification in the experiment above described the connecting conductor acquires certain properties which will be considered under the head of the Electrical Current.

(3) If a positively electrified body be brought near an insulated conductor, the latter shows signs of electrification. The end nearer the first body is negatively, the farther end positively, electrified. If the first body be removed, all signs of electrification on the conductor disappear. If, before the first body is removed, the conductor be joined to ground, the positive electrification disappears. If now the connection with ground be broken, and the first body removed, the conductor is negatively electrified.

The experiment can be carried out so as to give quantitative results, in a way first given by Faraday. An electrified body, for example a brass ball suspended by a silk thread, is introduced into the interior of an insulated closed metallic vessel. The exterior of the vessel is then found to be electrified in the same way as the ball. This electrification disappears if the ball be removed. If the ball be touched to the interior of the vessel, no change in the amount of the external electrification can be detected. If, after the ball is introduced into the interior, the vessel be joined to ground by a wire, all external electrification disappears. If the ground connection be broken, and the ball removed, the vessel has an electrification dissimilar to that of the ball. If the ball, after the ground connection is broken, be first touched to the interior of the vessel and then removed, neither the ball nor the vessel is any longer electrified.

A body thus electrified without contact with any charged body is said to be electrified by *induction*. The above-mentioned facts

show that an insulated conductor, electrified by induction, is electrified both positively and negatively at once; that the electrification of a dissimilar kind to that of the inducing body persists, however the insulation of the conductor be afterwards modified; and that the total positive electrification induced by a positively charged body is equal to that of the inducing body, while the negative electrification can exactly neutralize the positive electrification of the inducing body.

The use of the terms positive and negative is thus justified, since they express the fact that equal electrifications of dissimilar kinds are exactly complementary, so that if they be superposed on a body that body is not electrified. These two kinds of electrification may then be spoken of as opposite.

If the glass and resin considered in the first experiment be rubbed together within the vessel, and in general if any apparatus which produces electrification be in operation within the vessel, no signs of any external electrification can be detected. It is thus shown that, whenever one kind of electrification is produced, an equal electrification of the opposite kind is also produced at the same time.

Franklin showed that, by the use of a closed conducting vessel of the kind just described, a charged conductor introduced into its interior and brought into conducting contact with its walls is always completely discharged, and the charge is transferred to the exterior of the vessel. This procedure furnishes a method of adding together the charges on any number of conductors, whether they be charged positively or negatively. It is thus theoretically possible to increase the charge of such a conductor indefinitely.

(4) If any instrument for detecting forces due to electrifications be introduced into the interior of a closed conductor charged in any manner, it is found that no signs of force due to the charge can be detected. The experiment was accurately executed by Cavendish, and afterwards tried on a large scale by Faraday. It proves that within a closed electrified conductor there is no elec-

trical force due to the charge on the conductor, or that the potential due to the electrical forces is uniform within the conductor.

253. Law of Electrical Force.—If two charged bodies be considered, of dimensions so small that they may be neglected in comparison with the distance between the bodies, the stress between the two bodies due to electrical force is proportional directly to the product of the charges which they contain, and inversely to the square of the distance between them.

If Q and Q' represent two similar charges, r the distance between them, and k a factor depending on the units in which the charges are measured, the formula expressing the repulsion between them is $k\frac{QQ'}{r^2}$.

Coulomb used the torsion balance (§ 109) to demonstrate this law. At one end of a glass rod suspended from the torsion wire and turning in the horizontal plane is placed a gilded pith ball, and through the lid of the case containing the apparatus can be introduced a similar insulated ball, so arranged that its centre is at the same distance from the axis of rotation of the suspended system, and in the same horizontal plane, as the centre of the first ball. This second ball may be called the *carrier*.

To prove the law as respects quantities, the suspended ball is brought into equilibrium at the point afterwards to be occupied by the carrier ball. The carrier ball is then charged and introduced into the case. When it comes in contact with the suspended ball, it shares its charge with it and a repulsion ensues. The torsion head must then be rotated until the suspended ball is brought to some fixed point, at a distance from the carrier which is less than that which would separate the two balls in the second part of the experiment if no torsion were brought upon the wire. The repulsion is then measured in terms of the torsion of the wire. The charge on the carrier is then halved, by touching it with a third similar insulated ball, and, the charge on the suspended ball remaining the same, the repulsion between the two balls at the same distance is again observed. If the case be so large that no disturbing effect

of the walls enters, and if the balls be small and so far apart that their inductive action on one another may be neglected, the repulsion in the second case is found to be one half that in the first case. In general the problem is a far more difficult one, for the distribution on the two spheres is not uniform. That portion of the distribution dependent on the induction of the balls can be calculated, but the irregularities of distribution due to the action of the walls of the case and other disturbing elements can only be allowed for approximately.

The law as respects distance is proved in a somewhat similar way. The repulsions at two different distances are measured in terms of the torsion of the wire, the charges on the two balls remaining the same. The same corrections must be introduced as in the former case.

254. Distribution.—The law of electrical force has been stated in terms of the charges of two bodies. We may, however, consider electricity as a quantity which has an existence independent of matter and which is distributed in space. The fact cited in § 252 (4) shows that this distribution must be looked on as being on the surfaces of conductors, and not in their interiors. If we define *surface density* of electrification at any point on the surface of a charged conductor as the limit of the ratio of the quantity of electricity on an element of the surface at that point to the area of the element as that area approaches zero, we may measure quantities of electricity in terms of surface density. The surface density of electricity is usually designated by σ.

If the law of electrical force hold true not only for charges on bodies, but also for quantities of electricity on the surface elements of a conductor, it is evident, from the fact that within an electrified conductor there is no electrical force, that its surface density of electrification must be proportional at every point on its surface to the thickness at that point of a shell of matter which is so distributed on that surface that there is no force at any point enclosed by the surface. The distribution on a charged sphere may, from symmetry, be assumed uniform. The fact that there is no electri-

cal force within a charged sphere is then, from § 57, consistent with the law of electrical force which has been given; and since the means of detecting electrical force, if there were any, within a charged conductor are very delicate, this fact affords a strong corroborative proof of the law.

The determination of the distribution of electricity on irregularly shaped conductors is in general beyond our power. If we consider, however, a conductor in the form of an elongated egg, it can be readily seen that, in order that there may be no electrical force within it, the surface density at the pointed end must be greater than that anywhere else on its surface. In general, the surface density at points on a conducting surface depends upon the curvature of the surface, being greater where the curvature is greater. Thus, if the conductor be a long rod terminating in a point, the surface density at the pointed end is much greater than that anywhere else on the rod.

255. Unit Charge.—The law of electrical force enables us to define a *unit charge*, based upon the fundamental mechanical units.

Let there be two equal and similar positive charges concentrated at points unit distance apart in air, such that the repulsion between them equals the unit of force. Then each of the charges is a *unit charge*, or a *unit quantity of electricity*. With this definition of unit charge, it may be said that the force between two charges is not merely proportional to, but equals, the product of the charges divided by the square of the distance between them. The factor k in the expression for the force between two charges becomes unity, and the dimensions of $\frac{QQ'}{r^2}$ are those of a force. If the charges be equal, we have $\left[\frac{Q^2}{r^2}\right] = MLT^{-2}$. Hence $[Q] = M^{\frac{1}{2}}L^{\frac{3}{2}}T^{-1}$ are the dimensions of the charge. This equation gives the charge in absolute mechanical units, and by means of it all other electrical quantities may be expressed in absolute units. It is at the basis of the *electrostatic system* of electrical measurements.

The practical unit of charge or quantity is called the *coulomb*.

It is the quantity of electricity transferred during one second by a current of one ampere (§ 291).

256. Electrical Potential.—The electrical forces have a potential similar to that discussed in §§ 53–57. The unit quantity of positive electricity is taken as the test unit. Since (§ 252, (4)) the potential at every point of a charged conductor is the same, the surface of the conductor is an equipotential surface. The potential of this surface is often called the *potential of the conductor*. A conductor joined to ground is at the potential of the earth. It will be shown (§ 260) that the potential of the earth is not appreciably modified when a charged conductor is joined to ground.

For these reasons it is usual to take the potential of the earth as the fixed potential or zero from which to reckon the potentials of electrified bodies. The potential of a freely electrified conductor and of the region about it is thus positive when the charge of the conductor is positive, and negative when it is negative. A conductor joined to ground is at zero potential.

The difference of potential between two points is equal to the work done in carrying a unit quantity of electricity from one point to the other. We then have the equation $Q(V' - V) = \text{work}$. Hence follows the dimensional equation $[V' - V] = \dfrac{ML^2T^{-2}}{M^{\frac{1}{2}}L^{\frac{3}{2}}T^{-1}} = M^{\frac{1}{2}}L^{\frac{1}{2}}T^{-1}$, the dimensions of difference of potential in electrostatic units.

If any distribution of a charge exist on a conductor, which is such that the potential at all points in the conductor is not the same, it is unstable, and a rearrangement goes on until the potential becomes everywhere the same. The process of rearrangement is said to consist in a flow of electricity from points of higher to points of lower potential.

On this property of electricity depends the fact that a closed conducting surface completely screens bodies within it from the action of external electrical forces. For, whatever changes in potential occur in the region outside the closed conductor, a redistribution will take place in it such as to make the potential of every

point within it the same. Electrical force depends on the space rate of change of potential, and not on its absolute value. Hence the changes without the closed conductor will have no effect on bodies within it. Further, any electrical operations whatever within the closed conductor will not change the potential of points outside it. For, whatever operations go on, equal amounts of positive and negative electricity always exist within the conductor, and hence the potential of the conductor remains unaltered. Hence electrical experiments performed within a closed room yield results which are as valid as if the experiments were performed in free space.

The advantage gained by the use of the idea of potential in discussions of electrical phenomena may be illustrated by a statement of the process of charging a conductor by induction described in § 252 (3). To fix our ideas, let us suppose that the field of force is due to a positively electrified sphere, and that the body to be charged is a long cylinder. When this cylinder, previously in contact with the earth and therefore at zero potential, is brought end on to a point near the sphere, it is in a region of positive potential, and is itself at a positive potential. If we consider the original potentials at the points in the region now occupied by the cylinder, it is easily seen that the potential of points nearer the sphere was higher than that of those more remote. When the cylinder is brought into the field, therefore, the portion nearer the sphere is temporarily raised to a higher potential than the portion more remote. The difference of potential between these portions is annulled by a flow of electricity from the points of higher potential to those of lower potential at a rate depending on the conductivity of the cylinder. The end of the cylinder nearer the sphere is negatively charged, the end more remote is positively charged, and the two charged portions are separated by a line on the surface, called the neutral line, on which there is no charge.

If the cylinder be now joined to ground, a flow of electricity takes place through the ground connection, and it is brought to zero potential. The potential of the cylinder is therefore everywhere lower than the original potentials of the points in the region

which it occupies. This necessitates a negative charge distributed over the whole cylinder. In other words, the earth and the cylinder may be considered as forming one conductor charged by induction, in which the neutral line is not within the cylinder.

If the ground connection be broken the electrical relations are not disturbed. If the cylinder be now removed to a region of lower potential against the attraction of the sphere, work will be done against electrical forces, which reappears as electrical energy. The potential of the cylinder is lowered, and, if it be again connected with the earth, work will be done by a flow of electricity to it.

In § 57 it was shown that the forces on the opposite side of a sheet, in which the surface density is σ, differ by $4\pi\sigma$. Now the force within an electrified conductor vanishes, so that the force at a point just outside it is given by $4\pi\sigma$.

The pressure outwards on the surface of an electrified conductor due to the repulsion of the various parts of the charge for one another is equal to $2\pi\sigma^2$. For the force just outside the conductor, which is equal to $4\pi\sigma$, is due to that part of the conductor immediately under the point considered, which may be considered plane, and to the rest of the conductor. The force due to the plane part is (§ 57) equal to $2\pi\sigma$, and that due to the rest of the conductor is therefore also $2\pi\sigma$. Select any small portion of the surface of the conductor of area a. The force on unit quantity acting outward from the conductor at a point in that area due to the charge of the rest of the conductor is $2\pi\sigma$. This force acts on every unit of charge on the area. The force on the area acting outwards is then $2\pi a\sigma^2$, or the pressure at a point in the area referred to unit of area is $2\pi\sigma^2$. This quantity is often called the *electric pressure.*

257. Capacity.—The electrical *capacity* of a conductor is defined to be the charge which a conductor must receive to raise it from zero to unit potential, while all other conductors in the field are kept at zero potential. (~~This charge~~ varies for any one conductor in a way which cannot be always definitely determined, depending upon the medium in which the conductor is immersed and the

position of other conductors in the field.) When the charged conductor is in very close proximity to another conductor which is kept at zero potential, the amount of charge needed to raise it to unit potential is very great as compared with that required when the other conductor is more remote. Such an arrangement is called a *condenser*. If the charge on a conductor be increased, the increase in potential is directly as that of the charge. Hence the capacity C is obtained by dividing any given charge on a conductor by the potential of the conductor, or

$$C = \frac{Q}{V}. \tag{89}$$

The practical unit of capacity is the *farad*, which is the capacity of a conductor, the charge on which is one coulomb (§ 255) when its potential is one volt (§ 303). This unit is too great for convenient use. Instead of it a *microfarad*, or the one-millionth part of a farad, is usually employed.

The equation gives the dimensions of capacity. Measured in electrostatic units, they are $[C] = \left[\frac{Q}{V}\right] = \frac{M^{\frac{1}{2}}L^{\frac{3}{2}}T^{-1}}{M^{\frac{1}{2}}L^{\frac{1}{2}}T^{-1}} = L.$

258. Specific Inductive Capacity.—The capacity of a condenser of given dimensions depends upon the insulating medium used to separate its parts, or the *dielectric*. This was first discovered by Cavendish, and afterwards rediscovered by Faraday. If Q represent the charge required to raise a condenser in which the dielectric is vacuum to a potential V, then if another dielectric be substituted for vacuum, it is found that a different charge Q' is required to raise the potential to V. The ratio $\frac{Q'}{Q} = K$ is called the *specific inductive capacity*, or *dielectric constant*. Since $C' = \frac{Q'}{V}$ and $C = \frac{Q}{V}$ are the capacities of the condenser with the two dielectrics, it follows that

$$C' = CK, \tag{90}$$

where C is the capacity with vacuum as the dielectric. The specific

inductive capacity K is always greater than unity. Some dielectrics, such as glass and hard rubber, have a high specific inductive capacity, and at the same time are capable of resisting the strain put upon them by the electric stress to a much greater extent than such dielectrics as air. They are therefore used as dielectrics in the construction of condensers.

259. Condensers.—The simplest condenser, one which admits of the direct calculation of its capacity, and from which the capacities of many other condensers may be approximately calculated or inferred, consists of a conducting sphere surrounded by another hollow concentric conducting sphere which is kept always at zero potential by a ground connection. For convenience we assume the specific inductive capacity of the dielectric separating the spheres to be unity. Let the radius of the small sphere (Fig. 79) be denoted by R, that of the inner spherical surface of the larger one by R'; let a charge Q be given to the inner sphere by means of a conducting wire passing through an opening in the outer sphere, which may be so small as to be negligible. This charge Q will induce on the outer sphere an equal and opposite charge, $-Q$. Since the distribution on the surface of the spheres may be assumed uniform, the potential at the centre of the two spheres, due to the charge on the inner one, is $\frac{Q}{R}$, and the potential due to the charge of the outer sphere is $-\frac{Q}{R'}$. Hence the actual potential V at the centre, due to both charges, is $\frac{Q}{R} - \frac{Q}{R'} = Q\left(\frac{R' - R}{RR'}\right)$.

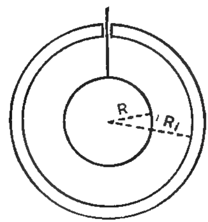

FIG. 79.

Hence the capacity is

$$C = \frac{Q}{V} = \frac{RR'}{R' - R}. \tag{91}$$

In order to find the effect of a variation of the value of R,

divide numerator and denominator by R' and write $C = \dfrac{R}{1 - \dfrac{R}{R'}}$.

Now, if R' be greater than R by an infinitesimal, the fraction $\dfrac{R}{R'}$, is less than unity by an infinitesimal, and the capacity of the accumulator is infinitely great. It becomes infinitely small if R be diminished without limit. The presence of any finite charge at a point would require an infinite potential at that point, which is of course impossible. The existence of finite charges concentrated at points, which we have assumed sometimes in order to more conveniently state certain laws, is therefore purely imaginary. If electricity is distributed in space, it is distributed like a fluid, a finite quantity of which never exists at a point.

If R' increase without limit, C becomes more and more nearly equal to R. Suppose the inner sphere to be surrounded not by the outer sphere but by conductors disposed at unequal distances, the nearest of which is still at a distance R' so great that $\dfrac{R}{R'}$, may be neglected in comparison with unity. Then if the nearest conductor were a portion of a sphere of radius R' concentric with the inner sphere, the capacity of the inner sphere would be approximately R. And this capacity is evidently not less than that which would be due to any arrangement of conductors at distances more remote than R'. Therefore the capacity of a sphere removed from other conductors by distances very great in comparison with the radius of the sphere is equal to its radius R. This value R is often called the capacity of a *freely electrified* sphere. Strictly speaking, a freely electrified conductor cannot exist; the term is, however, a convenient one to represent a conductor remote from all other conductors.

A common form of condenser consists of two flat conducting disks of equal area, placed parallel and opposite one another. The capacity of such a condenser may be calculated from the capacity of the spherical condenser already discussed. Let d represent the

distance $R' - R$ between the two spherical surfaces. Let A and A' represent the area of the surfaces of the two spheres of radius R and R'. Then we have $R^2 = \frac{A}{4\pi}$ and $R'^2 = \frac{A'}{4\pi}$. The capacity of the spherical condenser may then be written $\frac{\sqrt{AA'}}{4\pi d}$. If R' and R increase indefinitely, in such a manner that $R' - R$ always equals d, in the limit the surfaces become plane and A becomes equal to A'. The capacity therefore equals $\frac{A}{4\pi d}$. Since the charge is uniformly distributed, the capacity of any portion of the surface cut out of the sphere is proportional to the area S of that surface, or

$$C = \frac{S}{4\pi d}. \tag{92}$$

This value is obtained on the assumption that the distribution over the whole disk is uniform, and the irregular distribution at the edges of the disk is neglected. It is therefore only an approximation to the true capacity of such a condenser.

The so-called *Leyden jar* is the most usual form of condenser in practical use. It is a glass jar coated with tin-foil within and without, up to a short distance from the opening. Through the stopper of the jar is passed a metallic rod furnished with a knob on the outside and in conducting contact with the inner coating of the jar. To charge the jar, the outer coating is put in conducting contact with the ground, and the knob brought in contact with some source of electrification. It is discharged when the two coatings are brought in conducting contact. When the wall of the jar is very thin in comparison with the diameter and with the height of the tin-foil coating, the capacity of the jar may be inferred from the preceding propositions. It is approximately proportional directly to the coated surface, to the specific inductive capacity of the glass, and inversely to the thickness of the wall.

260. Systems of Conductors.—If the capacities and potentials of two or more conductors be known, the potential of the system

formed by joining them together by conductors is easily found. It is assumed that the connecting conductors are fine wires, the capacities of which may be neglected. Then the charges of the respective bodies may be represented by $C_1 V_1$, $C_2 V_2$, . . . $C_n V_n$, and the capacity of the system by the sum $C_1 + C_2 + \ldots C_n$. Hence V, the potential after connections have been made, is

$$V = \frac{C_1 V_1 + C_2 V_2 + \ldots C_n V_n}{C_1 + C_2 + \ldots C_n}.$$

In the case of two freely electrified bodies joined up together by a fine wire, we have $V = \dfrac{C_1 V_1 + C_2 V_2}{C_1 + C_2}$. When C_1 is very great compared with C_2, we obtain $V = V_1 + \dfrac{C_2}{C_1} V_2$.

Unless V_2 is so great that the term $\dfrac{C_2}{C_1} V_2$ becomes appreciable, the potential of the system is appreciably equal to the original potential of the larger body. The capacity of the earth, being equal to its radius, is very great in comparison with the capacity of any body used in our experiments, and hence the potential of the earth is not changed when it is connected with a charged body. This proposition justifies the adoption of the potential of the earth as the standard or zero potential.

261. Electroscopes and Electrometers.—An *electroscope* is an instrument used to detect the existence of a difference of electrical potential. It may also give indications of the amount of difference. It consists of an arrangement of some light body or bodies, such as a pith ball suspended by a silk thread, or a pair of parallel strips of gold-foil, which may be brought near or in contact with the body to be tested. The movements of the light bodies indicate the existence, nature, and to some extent the amount of the potential difference between the body tested and surrounding bodies.

An *electrometer* is an apparatus which gives precise measurements of differences of potential. The most important form is the *absolute* or *attracted disk electrometer*, originally devised by Harris,

and improved by Thomson. The essential portions of the instrument (Fig. 80) are a large flat disk B, which can be put in conducting contact with one of the two bodies between which the difference of potential is desired; a similar disk C, in the centre of which is cut a circular opening, placed parallel to and a little distance above the former one; a smaller disk A with a diameter a little less than that of the opening, which can be placed accurately in the opening and brought plane with the larger disk; and an arrangement, either a balance arm or a spring of known strength, from which the small disk is suspended, and by means of which the force acting on the disk when it is plane with the surface of the larger disk can be measured. The three disks can be conveniently styled the *attracting disk*, the *guard ring*, and the *attracted disk*. The position of the attracted disk when it is in the plane of the guard ring is often called the *sighted position*. The guard ring is employed in order that the distribution on the attracted disk may be uniform.

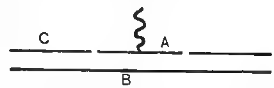

FIG. 80.

To determine the difference of potential between the attracted and attracting disks, we consider them first as forming a flat condenser. If we represent by Q the quantity of electricity on the attracted disk, by V and V_1 the potentials of the attracted and attracting disks respectively, by d the distance between them, and by S the area of the attracted disk, then, as has been shown in § 259, the capacity of such a condenser is $\frac{Q}{V_1 - V} = \frac{S}{4\pi d}$. Now from the nature of the condenser, and in consequence of the regular distribution due to the presence of the guard ring, we have $\frac{Q}{S} = \sigma$, the surface density on either plate, whence $\sigma = \frac{V_1 - V}{4\pi d}$. The surface density cannot be measured, and must be eliminated by means of an equation obtained by observation of the force with which the two disks are attracted. The plates are never far apart, and the force on a unit charge due to the charge on the lower one may be

always taken in the space between the plates as equal to $2\pi\sigma$ (§ 57). Every unit on the attracted disk is attracted with this force, and the total attraction, which is measured by means of the balance or spring, is $F = 2\pi\sigma^2 S$. Substituting this value of σ in the former equation, we get

$$V_1 - V = d\sqrt{\frac{8\pi F}{S}}, \tag{93}$$

which gives the difference of potential between the two plates in terms which are all measurable in absolute units.

Thomson's *quadrant electrometer* is an instrument which is not used for absolute measurement, but being extremely sensitive to minute differences of potential, it enables us to compare them with each other and with some known standard. The construction of the apparatus can best be understood from Fig. 81. Of the four metallic quadrants which are mounted on insulating supports, the two marked P and the two marked N are respectively in conducting contact by means of wires. The body C, technically called the *needle*, is a thin sheet of metal, suspended symmetrically just above the quadrants by two parallel silk fibres, forming what is known as a bifilar suspension. When there is no charge in the apparatus, the axes of symmetry of the needle lie above the spaces which separate the quadrants.

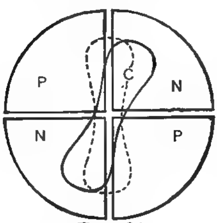

Fig. 81.

To use the apparatus, the needle is maintained at a high, constant potential, and the two points, the difference of potential between which is desired, are joined to the pairs of quadrants P and N. The needle is deflected from its normal position, and the amount of deflection is an indication of the difference of potential between the two pairs of quadrants.

262. Electrical Machines.—*Electrical machines* may be divided into two classes: those which depend for their operation upon friction, and those which depend upon induction.

The *frictional machine*, in one of its forms, consists of a circu-

lar glass *plate*, mounted so that it can be turned about an axis, and a *rubber* of leather, coated with a metal amalgam, pressed against it. The rubber is mounted on an insulating support, but, during the operation of the machine, it is usually joined to ground. Diametrically opposite is placed a row of metal points, fixed in a metallic support, constituting what is technically called the *comb*. The comb is usually joined to an accessory part of the machine presenting an extended metallic surface, called the *prime conductor*. The prime conductor is carried on an insulating support.

When the plate is turned, an electrical separation is produced by the friction of the rubber, and the rubbed portion of the plate is charged positively. When the charged portion of the plate passes before the comb, an electrical separation occurs in the prime conductor due to the inductive action of the plate, a negative charge passes from the comb to neutralize the positive charge of the plate, and the prime conductor is charged positively. Since accessions are received to the charge of the prime conductor as each portion of the plate passes the comb, it is evident that the potential of the prime conductor will continuously rise, until it is the same as that of the plate, or until a discharge takes place.

The fundamental operations of all *induction machines* are presented by the action of the *electrophorus*, an instrument invented by Volta in 1771. It consists of a plate of sulphur or rubber, which rests on a metallic plate, and a metallic disk mounted on an insulating handle. The sulphur is electrified negatively by friction, and the disk, placed upon it and joined to ground, is charged positively by induction. When the ground connection is broken and the disk lifted from the sulphur, its positive charge becomes available. The process is precisely similar to that described in § 256. It may evidently be repeated indefinitely, and the electrophorus may be used as a permanent source of electricity.

It is evident that a charged metallic plate may be substituted for the sulphur in the construction of an electrophorus, provided that the disk be not brought in contact with it, but only near it. A plan by which this is realized, and at the same time an imper-

ceptible charge on one plate is made to develop an indefinite quantity of electricity of high potential, is shown in Fig. 82. A_1 and A_2 are conducting plates, called *inductors*. In front of them two disks B_1 and B_2, called *carriers*, are mounted on an arm so as to turn about the axis E. Projecting springs b_1 and b_2 attached to these disks are so fixed as to touch successively the pins D_1 and D_2, connected with the plates A_1 and A_2, and the pins C_1 and C_2, insulated from the plates, but joined to the prime conductors F_1 and F_2.

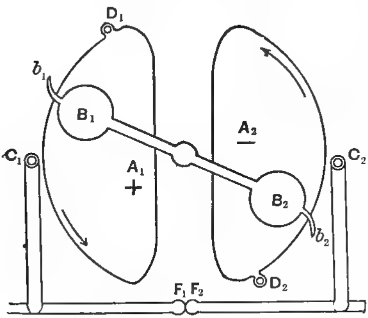

FIG. 82.

Suppose the prime conductors to be in contact and the carriers so placed that B_1 is between D_1 and C_1, and suppose the plate A_1 to be at a slightly higher potential than the rest of the machine. The carrier B_1 is then charged by induction. When the carriers are turned in the direction of the arrows, and the carrier B_1 makes contact with the pin C_1 it loses a part of its positive charge and the prime conductors become positively charged. At the same time the carrier B_2 becomes positively charged. As the carrier B_2 passes over the upper part of the plate A_2, the lower part of the plate A_2 is charged positively by induction. This positive charge is neutralized by the negative charge of the carrier B_1, when contact is made at D_2. The plate A_2 is then negatively charged. The carrier B_2 at its contact at D_1 shares its positive charge with the plate A_1. The carriers then return to the positions from which they started, and the difference of potential between the plates A_1 and A_2 is greater than it was at first. When, after sufficient repetition of this process, the difference of potential has become sufficiently great, the prime conductors may be separated, and the transfer of electricity between the points F_1 and F_2 then takes place through the air. Obviously the number of carriers may be increased, with a corresponding increase in the rapidity of

action of the machine. This improvement is usually effected by attaching disks of tin-foil at equal distances from each other on one face of a glass wheel, so that, as the wheel revolves, they pass the contact points in succession.

Another induction machine, invented by Holtz, differs in plan from the one just described in that the metallic carriers are replaced by a revolving glass plate, and the two metallic inductor plates by a fixed glass plate. In the fixed plate are cut two openings, diametrically opposite. Near these openings, and placed symmetrically with respect to them, are fixed upon the back of the plate two paper sectors or *armatures,* terminating in points which project into the openings. In front of the revolving plate and opposite the ends of the armatures nearest the openings are the combs of two prime conductors. Opposite the other ends of the armatures, and also in front of the revolving wheel, are two other combs joined together by a cross-bar.

In order to set this machine in operation, one of the paper armatures must be charged from some outside source. The surface of the revolving plate performs the functions of the carriers in the induction machine already explained. The armatures take the place of the inductors, and the points in which they terminate serve the same purpose as the contact points in connection with the inductors. The explanation of the action of this machine is, in general, similar to that already given. The effect of the combs joined by the cross-bar is equivalent to joining to ground that portion of the outside face of the revolving plate which is passing under them.

263. Energy of a System of Charged Bodies.—If the charge on a body be changed, the potential at every point in the field changes in the same proportion. To obtain the energy of a system of charged bodies we may apply the method used in § 248 to obtain the energy of a system of magnets. If Q represent the charge of one of the bodies and V its potential, the energy of the system is given by $\frac{1}{2}\Sigma QV$.

264. Strain in the Dielectric.—An instructive experiment illus-

trating Faraday's theory that the electrification of a conductor is due to an action in the dielectric surrounding it, may be performed with a jar so constructed that both coatings can be removed from it. If the jar be charged, the coatings removed by insulating handles without discharging the jar, and examined, they will be found to be almost without charge. If they be replaced, the jar will be found to be charged as before. The jar will also be found to be charged if new coatings similar to those removed be put in their place. This result shows that the true seat of the charge is in the dielectric. The experiment is due to Franklin.

That the arrangement in the dielectric is of the nature of a strain is rendered probable by the fact, first noticed by Volta, that the volume occupied by a Leyden jar increases slightly when the jar is charged. Similar changes of volume were observed by Quincke in fluid dielectrics as well as in different solids.

Another proof of the strained condition of dielectrics is found in their optical relations. It was discovered by Kerr that dielectrics previously homogeneous become doubly refracting when subjected to a powerful electrical stress. Maxwell has shown, from the assumptions of his electromagnetic theory of light, that the index of refraction of a transparent dielectric should be proportional to the square root of its specific inductive capacity. Numerous experiments, among which those of Boltzmann on the index of refraction of light in gases and those of Hertz and others on the index of refraction of electromagnetic waves in solids and liquids are the most striking, show that this predicted relation is very close to the truth.

It has further been shown that the specific inductive capacity of sulphur has different values along its three crystallographic axes. This is probably true also for other crystals.

Some crystals, while being warmed, exhibit on their faces positive and negative electrifications, which are reversed as the crystals are cooling. This fact, while as yet unexplained, is probably due to temporary modifications of molecular arrangement by heat.

If a jar be discharged and allowed to stand for a while, a second

discharge can be obtained from it. By similar treatment several such discharges can be obtained in succession. The charge which the jar possesses after the first discharge is called the *residual charge.* It does not attain its maximum immediately, but gradually, after the first discharge. The attainment of the maximum is hastened by tapping on the wall of the jar. This phenomenon was ascribed by Faraday to an absorption of electricity by the dielectric, but this explanation is at variance with Faraday's own theory of electrification. Maxwell explains it by assuming that want of homogeneity in the dielectric admits of the production of induced electrifications at the surfaces of separation between the non-homogeneous portions. When the jar is discharged the induced electrifications within the dielectric tend to reunite, but, owing to the want of conductivity in the dielectric, the reunion is gradual. After a sufficient time has elapsed, the alteration of the electrical state of the dielectric has proceeded so far as to sensibly modify the field outside the dielectric. The residual charge then appears in the jar. This explanation is supported by the fact that no residual charge remains when the dielectric is a fluid.

265. Tubes of Electrical Force.—If we admit that the nature and condition of the dielectric between conductors determine the charge upon them, an admission which the facts of specific inductive capacity and those cited in the last section render necessary, we must conclude that the hypothesis of electrical charges acting on each other directly at a distance, which we have used up to this point, is an artificial one, and that a more accurate representation of the real state of an electrical field will be had by assuming the action between the electrified bodies to be due to an action in the dielectric. We cannot explain the relation between electricity and the condition of the dielectric which will cause the actions observed between the electrified bodies, but we can show that these actions are consistent with certain conditions in the dielectric which are mechanically possible.

Let us consider a positively charged conductor A which is everywhere surrounded with other conductors. We may assume

that these other conductors are at any distance from A, and that they are at the common potential zero. They are then equivalent to a single conductor B surrounding the conductor A. Lines of force start from every point of A and pass to corresponding points of B. Mark out a small area on the surface A, and consider the closed surface formed by the lines of force passing through the contour of that area and surfaces drawn in the dielectric just outside the conductor A and just outside the conductor B. This closed surface is a tube of force, and if F_A and F_B represent the forces acting at the two cross-sections of the tube at A and B respectively, and s_A and s_B represent the areas of those cross-sections, we have (§ 56), $F_A s_A = - F_B s_B$, the forces being considered as directed along the normals drawn outward from the conductors. Since the force within the conductors vanishes, the force just outside the surface of A is $F_A = 4\pi\sigma_A$, and that just outside the surface of B is $F_B = 4\pi\sigma_B$. Using these values for F_A and F_B, we have $\sigma_A s_A = - \sigma_B s_B$. Now these products are equal to the quantities of electricity present on the areas s_A and s_B, so that we have $q_A = - q_B$. The charges at the two ends of the tube of force are therefore equal and of opposite sign. Since the tubes of force which proceed from A either extend to infinity or end on conductors, the charges on those conductors are never greater than the total charge on A. If, as we have assumed, the conductors B completely surround A, the charges on B are equal to the charge on A. If we divide the surface of A into areas upon each of which a unit charge of electricity is present, and erect tubes of force upon those areas, the dielectric will be mapped out by those tubes. Such a tube may be called a *unit tube* or a *Faraday tube,* in accordance with the proposition of J. J. Thomson.

266. Electrical Forces explained by Tubes of Force.—The strength of the field at any point in the dielectric is inversely as the area of the normal cross-section of the unit tube of force at that point. For, by § 56, the product Fs is constant throughout the tube. At the surface of the conductor from which the tube starts, F is equal to $4\pi\sigma$ and $Fs = 4\pi\sigma s = 4\pi$, since s is the cross-

section of the unit tube at the charged surface, and σs is therefore equal to unity. The force F at any point in the dielectric is therefore equal to $\frac{4\pi}{s}$, or is inversely as the cross-section of the tube. If we represent by N the number of unit tubes which pass through a unit area of an equipotential surface, and if we assume that the force is appreciably constant over this area, we have $N = \frac{1}{s}$ and $F = 4\pi N$, that is, the force at any point in the field is proportional to the number of unit tubes which pass perpendicularly through a unit area at that point.

In the discussion up to this point we have assumed that the medium between the two conductors has the specific inductive capacity or dielectric constant unity. If the dielectric constant be not unity, but some other number, say K, the difference of potential between A and B that will be produced by a given charge on A is less than that which will be produced when the dielectric constant is unity, in the ratio of 1 to K. The general expression for the force in the field is therefore $F = \frac{4\pi N}{K}$.

The electric pressure or force on unit area of the surface of the conductor, when the conductor is surrounded by a medium of which the dielectric constant is K, is given by $\frac{2\pi\sigma^2}{K}$. This may be seen at once by applying the proof of § 256 to this case, remembering that, as has just been shown, the force outside the conductor is given by $\frac{4\pi\sigma}{K}$. Now on the view here taken, that the electrical forces are due to actions in the dielectric, this pressure should not be looked at as the result of the repulsions of the various elements of charge on the conductor, but rather as the result of some action in the dielectric. This action must be a pull or tension on the surface of the conductor. Since σ represents the number of unit tubes which proceed from unit area of the conductor, this pull is equal to $\frac{2\pi\sigma}{K}$ applied to the end of each unit tube, or since the force

just outside the surface is equal to $\frac{4\pi\sigma}{K}$, the pull on the end of each unit tube is also given by $\frac{F}{2}$.

The forces which act upon electrified bodies may therefore be considered as arising from tensions in the unit tubes, provided these tensions are not mechanically impossible. It may be shown that a medium in which such tensions exist is not in equilibrium unless pressures numerically equal to the tensions, and at right angles to them, act throughout the medium. In order, therefore, that we may adequately represent the electrical field by the aid of unit tubes, we must assume a tension in each of these tubes along the lines of force and a pressure in every direction at right angles to it of the same numerical value. The tensions tend to shorten the tubes, the pressures to repel them from one another. All the forces which act between electrified bodies may be explained in terms of these actions between the tubes of force.

267. Energy in the Dielectric.—The tension on the cross-section of the unit tube at any point in the field is also $\frac{F}{2}$, where F represents the force at that point. To show this, it is sufficient to suppose one of the equipotential surfaces around the charged body replaced by a conductor maintained at the potential of that surface. The distribution in the field between the two conductors will then be the same as before. By reasoning similar to that already employed, it is seen at once that the force on the surface of the new conductor which carries unit charge, or the pull on the end of a unit tube at that surface, is given by $\frac{F}{2}$, where F is the force at a point in the end of the unit tube. No restriction has been made as to the particular equipotential surface chosen to be replaced by a conductor, and thus it appears that the tension or pull on the cross-section of the tube of force is everywhere equal to $\frac{F}{2}$, where F is the force at a point in that cross-section.

To find the tension or pull across unit area normal to the lines

of force, we notice that if N represent the number of tubes of force which pass through unit area drawn at a certain point in the field normal to the lines of force, and if s represent a small area normal to the lines of force, then Ns represents the number of tubes of force which pass through that area, and the tension on that area is $\frac{1}{2}FNs$. Now we have seen that in any field in which the dielectric constant is K, $F = \frac{4\pi N}{K}$. Hence, substituting for N and dividing by s, we have the tension on unit area given by $\frac{2\pi N^2}{K} = \frac{KF^2}{8\pi}$.

On the view which we are now taking it is natural to consider the work done in charging bodies in a field as expended in modifying the dielectric or in setting up unit tubes in it. We will examine on this supposition the distribution of energy in the electrical field. It has already been proved (§ 263) that the energy of a system of charged bodies is equal to $\frac{1}{2}\Sigma QV$, where Q is the charge and V the potential of each body. Let us consider a tube of force starting from a body at potential V_1 and proceeding to another body at potential V_2; the charges at the ends of these tubes of force are equal and of opposite sign. The energy of the first conductor due to the portion of its charge we are now considering, which may be called q, is $\frac{1}{2}qV_1$; the energy of the second conductor due to the corresponding equal charge is $\frac{1}{2}qV_2$. The energy, therefore, due to the charges associated with a tube of force is $\frac{1}{2}q(V_1 - V_2)$. All charges in the field may be associated in this way in pairs, and the total energy of the field expressed by $\frac{1}{2}\Sigma q(V_1 - V_2)$. Now $V_1 - V_2$ measures the work done by the electrical forces in moving a unit charge from the first conductor to the second. If F represent the force at any point in a tube of force and Δl an element of length of the tube, the product $F\Delta l$ represents the work done in moving the unit charge along that element, and the total work done in moving over the length of the whole tube is $\Sigma F\Delta l = V_1 - V_2$. The energy associated with the whole tube is therefore $\frac{1}{2}q\Sigma F\Delta l$, and if we assume the unit length so small that the force does not appreciably vary within it, this energy may be considered as distributed

throughout the tube in such a way that each unit of length of the tube contains the energy $\frac{1}{2}qF$. If the tube be a unit tube so that $q = 1$, each unit of length of this tube will have in it a quantity of energy equal to $\frac{1}{2}F$.

To find the energy in unit volume of the dielectric, we consider a small cylinder, its height l being taken along the lines of force and its base s normal to them. The number of unit tubes which pass through the base is Ns. Since the energy in unit length of each of these tubes is $\frac{1}{2}F$, and since therefore the energy in the length l is $\frac{1}{2}Fl$, we have the energy in the volume ls equal to $\frac{1}{2}FNls$, or the energy in unit volume equal to $\frac{1}{2}FN$. Now we have $F = \frac{4\pi N}{K}$, so that the energy in unit volume is $\frac{2\pi N^2}{K} = \frac{KF^2}{8\pi}$.

By comparing this result with the value obtained for the tension across unit area it appears that the tension across unit area and the energy of unit volume are numerically equal. They both vary from point to point in the dielectric, depending upon the electrical force at each point. Unless the force is appreciably constant for all points of a finite region, the actual tension across a unit area and the actual energy of unit volume will not be given accurately by these expressions: they are more strictly the limits of the ratios between the tension and the area on which it acts, and the energy and the volume containing it.

268. Forces on Electrified Bodies.—It has already been stated that the stresses between charges may be represented by supposing that the tubes of force exert a tension along the lines of force and an equal pressure in all directions perpendicular to the lines of force, or as may be said, the lines of force tend to diminish in length and to repel each other. This mode of conceiving the stresses between charged bodies may be illustrated in some simple cases without the aid of diagrams of lines of force. The lines of force around a uniformly charged sphere are radial and the tubes of force are similar cones; if the sphere be charged positively, the force is directed outward from it, and if charged negatively, is

directed toward it. When two such spheres, one charged positively and the other negatively, are brought near each other, the tubes of force in the region between them to some extent coincide, so that the number of tubes of force which pass through unit area in the region between them is greater than that passing through the same area when only one of the spheres is present. On the other hand, the tubes in the region outside both the spheres counteract each other, and the number of tubes of force which pass through unit area in this region is less than when only one of the spheres is present. It may be seen thus roughly, and a diagram of the actual tubes of force in the field shows clearly, that the number of tubes of force which proceed from unit area of either one of the spheres on the surfaces confronting each other is greater than the number which proceeds from unit area on the outer surfaces. The tensions tending to draw the spheres together are thus greater than the tensions tending to separate them, and the spheres therefore appear to attract each other. If the two spheres which are brought near each other have similar charges, the tubes of force in the region between them are opposed to each other and the number of tubes of force in that region is therefore diminished, while in the region outside the two spheres their tubes of force partly coincide and the number is increased. The tension is therefore greater on the outer surfaces of the spheres, and they are pulled apart or appear to repel each other. In these explanations no account has been taken of the inductive action of one sphere on the other.

We may use the results obtained in the last section in the discussion of the forces which act upon a body originally uncharged, having a dielectric constant K and brought into an electrical field set up in a medium of which the dielectric constant is different from K; for convenience, we will assume it to be unity. Let us assume that the body to be brought into the field is small and represent its volume by v. Now, before the body is brought into the field the energy in the volume afterwards occupied by it is $2\pi N^2 v$. The energy in the same volume, after it is occupied by the body, is $\frac{2\pi N^2 v}{K}$. Now we know by experiment that K is always

greater than unity, so that the introduction of such a body into the field involves a loss of energy, and this loss of energy is greater as N is greater. Bodies tend to move so as to make their potential energy a minimum, and the given body will therefore move from a place of weaker to a place of stronger electrical force. This conclusion is reached on the supposition that the electrical field is not modified by the presence of the body—a supposition which can be made only when K is very nearly equal to unity. When K is not nearly equal to unity, the potential is not only diminished by the movement of the body from a place of weaker to a place of stronger electrical force, but also by the movement into it of the tubes of force; for a unit tube of force is associated with less energy in a medium of which the dielectric constant is K than in the medium of which the dielectric constant is unity, and the potential energy of the field is therefore diminished by a crowding of the tubes of force into the given body. This process cannot go on indefinitely so that the body includes all the tubes of force of the field, for as some of them enter the body others outside of it are lengthened and their energy is thereby increased. The concentration of the tubes in the body ceases, therefore, when the loss of energy due to their entrance into the body is balanced by the gain of energy due to the lengthening of those outside the body.

A conductor may be looked on as a body having a dielectric constant $K = \infty$. There is no electrical force within a conductor, and the energy lost by the field in consequence of a conductor being introduced into it is $\frac{F^2 v}{8\pi}$, where v is the volume of the conductor. This loss of energy is greater as F is greater, and the conductor therefore tends to move from a place of weaker to a place of stronger electrical force. There will also be a diminution of potential energy due to the concentration of tubes of force upon the conductor; the conductor disturbs the electrical field and concentrates the tubes of force upon it in a way similar to that of the body just described, but to a greater extent.

269. Cause of the Stress in the Dielectric.—The theory that the

electrical forces are due to stresses in the medium between the electrified bodies serves very well in expressing the results of experiment, but it gives no information about the origin or cause of the stresses in the medium. Faraday, who originated the theory, apparently thought that they arise from the electrification by induction of the separate particles or molecules of the medium in such a way that they resemble, so far as their external action is concerned, the magnetic molecules in Weber's theory of magnetism. This view was not held consistently even by its author, and cannot be accepted if we remember that electrical actions take place through vacuum. Maxwell conceived of electricity as distributed everywhere in space, and considered the charging of a body as a displacement of the electricity in the region around it in one sense if the body is charged positively, in the opposite sense if charged negatively. Conductors offer no resistance to such a displacement, but in dielectrics the displacement is resisted by an action which Maxwell called electrical elasticity. This mode of describing the electrical field is satisfactory so long as the field is considered in equilibrium, but becomes difficult of application when movements of charges occur in the field. J. J. Thomson has shown that all the phenomena of the electrical field may be described in terms of the motions or interactions of tubes of force, one of which is supposed to be connected with each atom of matter in the field. Thomson gives no mechanical explanation of the properties which these tubes must be assumed to have, only saying that "the analogies between their properties and those of the tubes of vortex motion irresistibly suggest that we should look to the rotary motion in the ether for their explanation."

CHAPTER III.

THE ELECTRICAL CURRENT.

270. Fundamental Effects of the Electrical Current.—In 1791 Galvani of Bologna published an account of some experiments made two years before, which opened a new department of electrical science. He showed that, if the lumbar nerves of a freshly skinned frog be touched by a strip of metal and the muscles of the hind leg by a strip of another metal, the leg is violently agitated when the two pieces of metal are brought in contact. Similar phenomena had been previously observed when sparks were passing from the conductor of an electrical machine in the vicinity of the frog preparation.

He ascribed the facts observed to a hypothetical animal electricity or vital principle, and discussed them from the physiological standpoint; and thus, although he and his immediate associates pursued his theory with great acuteness, they did not affect any marked advance along the true direction. Volta at Pavia followed up Galvani's discovery in a most masterly way. He showed that if two different metals, or, in general, two heterogeneous substances, be brought in contact, there immediately arises a difference of electrical potential between them. He divided all bodies into two classes. Those of the *first class*, comprising all simple bodies and many others, are so related to one another that, if a closed circuit be formed of them or any of them, the sum of all the differences of potential taken around the circuit in one direction is equal to zero. If a body of the *second class* be substituted for one of the first class, this statement is no longer true. There exists then in the circuit a preponderating difference of potential in one direction.

Volta described in 1800 an arrangement for utilizing these properties of bodies for the production of continuous electrical currents. He placed in a vessel, containing a solution of salt in water, plates of copper and zinc separated from one another. When wires joined to the copper and zinc were tested, they were found to be at different potentials, and they could be used to produce the effects observed by Galvani. The effects were heightened, and especially the difference of potential between the two terminal wires was increased, when several such cups were used, the copper of one being joined to the zinc of the next, so as to form a series. This arrangement was called by Volta the galvanic battery, but is now generally known as the *voltaic battery.*

Volta observed that if the terminals of his battery were joined the connecting wire became heated.

Soon after Volta sent an account of the invention of his battery to the Royal Society, Nicholson and Carlisle observed that, when the terminals of the battery were joined by a column of acidulated water, the water was decomposed into its constituents, hydrogen and oxygen.

In 1820 Oersted made the discovery of the relation between electricity and magnetism. He showed that a magnet brought near a wire joining the terminals of a battery is deflected, and tends to stand at right angles to the wire. His discovery was at once followed up by Ampère, who showed that, if the wire joining the terminals be so bent on itself as to form an almost closed circuit, and if the rest of the circuit be so disposed as to have no appreciable influence, the magnetic potential at any point outside the wire will be similar to that due to a magnetic shell.

In 1834 Peltier showed that, if the terminals of the battery be joined by wires of two different metals, there is a production or an absorption of heat at the point of contact of the wires, depending upon which of the wires is joined to the terminal the potential of which is positive with respect to the other. This fact is referred to as the *Peltier effect.*

271. The Electrical Current.—In 1833 Faraday showed con-

clusively that if a Leyden jar be discharged through a circuit, it will momentarily produce thermal, chemical, and magnetic effects which are similar to those just described as produced continuously by the voltaic battery.

The discharge of the jar may be variously represented. So long as electricity is considered as a fluid or substance, it is easiest to think of the discharge as the transfer of electricity from a place of higher to a place of lower potential, or rather as the equalization of potential by the transfer of equal and opposite quantities in opposite senses, and to explain the continuous effects produced by the voltaic battery by ascribing them to a *current*, or continuous transfer of electricity around the circuit. This view is capable of representing most of the phenomena of steady or permanent currents, but it is less successful in representing the phenomena of variable currents. If we consider electrical phenomena as due to actions in the dielectric, we may obtain a more adequate representation of the discharge and also of all the phenomena of the current by the use of the unit tubes of force described in § 265. We may obtain some idea of the connection of these tubes with the current if we examine their behavior during the discharge of a condenser.

To make the discussion as simple as possible, we suppose the condenser to be made of two equal plates A and B; their potentials are V_A and V_B, V_A being the greater. The lines of force originate at A and pass to B in the manner shown in Fig. 83. This figure has been roughly copied from the one given by J. J. Thomson. Let Q represent that part of the charge on A to which corresponds an equal and opposite charge on B: the number of unit tubes of force which pass from A to B will then be given by Q. Now let us join A to B by a conductor C, which for the sake of simplicity shall coincide with the direction of the lines of force. No tube of force can exist within a conductor, and those which were present in the volume which the conductor

FIG. 83.

occupies immediately disappear. For our purposes the manner of this disappearance of the tubes of force is of no consequence: it may be described as a shrinking of the tubes in the conductor, so that the ends which leave the plates of the condenser move toward the middle of the conductor and at last meet. The disappearance of these tubes from the conductor relieves those lying immediately around it of the lateral pressure which maintained them in equilibrium, and they are accordingly driven into the conductor and in turn disappear. This process is continued until all the tubes of force have disappeared. The discharge of the condenser may therefore be represented as the lateral movement of the tubes of force originally in the field and their disappearance in the conductor. The discharge is not really so simple as it is here supposed to be. We have supposed the process to cease when the difference of potential between the two plates becomes zero, but this is in fact not the case: the discharge is really oscillatory, the difference of potential being alternately positive and negative, rapidly diminishing in numerical value until it disappears. In order to account for this, something analogous to inertia must be ascribed to the tubes of force.

While the discharge is going on, a magnetic field exists around the conductor. If the discharge be thought of as being merely the transfer of charge along the conductor, there is no apparent mechanism connecting the discharge with the magnetic force, but on the view now being presented the magnetic field may be thought of as due in some way to the movement through the dielectric of the tubes of force. If the discharge pass through a compound body, capable of decomposition by it, a portion of that body will be resolved into two constituents. On the view that the discharge is a mere transfer of charge, these constituents must be supposed to serve as carriers of that charge, but this view cannot represent the connection between these constituents and their charges, nor the conditions which enable them to give up their charges. On the other hand, by supposing each unit of the constituent to be invariably associated with a tube of force, we may describe the discharge through such a chemical compound in terms of the changes which

take place in the tubes of force, in a manner consistent with what we know of their nature. Thus this latter view furnishes a more adequate representation of the discharge than the older and simpler view.

We have already seen (§ 267) that the energy contained in each unit tube is equal to one half the difference of potential between its ends. Since Q represents the number of unit tubes which pass between the plates, the energy of the field is $\frac{1}{2}Q(V_A - V_B)$; after the discharge this energy has entered the conductor.

If an arrangement be effected by which the difference of potential between A and B is kept constantly equal to $(V_A - V_B)$, the work done by the transfer of Q units of charge, and therefore the energy lost by the disappearance of Q unit tubes of force, is $Q(V_A - V_B)$. Let W represent the energy lost by such a continuous discharge or *current* in unit time, and t the time in which Q tubes of force disappear. Then $Wt = Q(V_A - V_B)$, and

$$W = \frac{Q}{t}(V_A - V_B). \tag{94}$$

The ratio $\frac{Q}{t}$ is represented by I and called the *current strength* or simply the *current* in the conductor. It may be variously considered as the rate of transfer of charge between the conductors, or as the rate at which the unit tubes of force are destroyed.

272. Electrostatic Unit of Current.—Let us denote the potentials at the two points 1 and 2 in a circuit by V_1 and V_2, and let V_1 be greater than V_2: then if, in the time t, a quantity of electricity equal to Q passes through a conductor joining those points from potential V_1 to potential V_2, the energy expended is $Q(V_1 - V_2)$.

If the conductor be a single homogeneous metal or some analogous substance, and if no motion of the conductor or of any external magnetic body take place, the energy expended in the conductor is transformed into heat. If we suppose this transformation to go on at a uniform rate, and denote the heat developed in unit time by H, we may substitute H for W in equation (94), and have

$$H = I(V_1 - V_2). \tag{95}$$

If heat and difference of potential be measured in absolute units, this equation enables us to determine the absolute *unit of current.* The system of units here used is the electrostatic system. The dimensions of current strength in the electrostatic system are obtained from this equation. We have $[I] = \left[\frac{Q}{t}\right] = M^{\frac{1}{2}}L^{\frac{3}{2}}T^{-2}$.

273. Electromotive Force.—We may consider the current as an operation by which energy is transformed in the conductor, either by the transfer of electricity through it or by the entrance of tubes of force into it and their disappearance within it. In the example before us, we have assumed that the transfer or the movement of the tubes was due to some cause which set up a difference of potential between parts of the conductor. This condition is not necessary for the maintenance of a current; in certain circumstances energy may be expended in a conductor, without the existence of a finite difference of potential between neighboring parts of the conductor. The power of establishing and sustaining the conditions which make a continuous expenditure of energy possible is called *electromotive force.*

The energy expended in unit time in a circuit carrying the current I, and in which the electromotive force is E, is

$$W = IE. \tag{96}$$

In a circuit containing a voltaic battery, if two points on the circuit outside the battery be tested by an electrometer, a difference of potential between them will be found. If the circuit be broken between the two points considered, the difference of potential between them becomes greater. This maximum difference of potential is the sum of finite differences of potential supposed to be due to molecular interactions at the surfaces of contact of different substances in the circuit, and is the measure of the electromotive force of the battery.

274. Ohm's Law.—In § 252 it was remarked that a body is distinguished as a good or a poor conductor by the rate at which it

will equalize the potentials of two electrified conductors, if it be used to connect them. Manifestly this property of the substances forming a circuit will influence the strength of the current in the circuit. It was shown on theoretical considerations, in 1827, by Ohm that in a homogeneous conductor which is kept constant the current varies directly with the difference of potential between the terminals. If R represent a factor, constant for each conductor, *Ohm's law* is expressed in its simplest form by

$$IR = V_1 - V_2. \tag{97}$$

The quantity R is called the *resistance* of the conductor. If the difference of potential be maintained constant, and the conductor be altered in any way that does not introduce an internal electromotive force, the current will vary with the changes in the conductor, and there will be a different value of R with each change in the conductor. The quantity R is therefore a function of the nature and materials of the conductor, and does not depend on the current or the difference of potential between the ends of the conductor. Since it is the ratio of the current to the difference of potential, and since we know these quantities in electrostatic units, we can measure R in electrostatic units. From the dimensions of I and $(V_1 - V_2)$ we may obtain the dimensions of R. They are in electrostatic units

$$[R] = \left[\frac{V_1 - V_2}{I}\right] = L^{-1}T.$$

Since the difference of potential in equation (97) is the measure of the electromotive force in the conductor considered, it is natural to extend the relation therein expressed to the whole circuit, in which the current is maintained by the electromotive force E. The expression of Ohm's law for the whole circuit is

$$IR = E. \tag{98}$$

This relation cannot in every case be experimentally verified, but in many cases in which the electromotive force may be directly and accurately calculated its validity has been demonstrated.

275. Specific Conductivity and Specific Resistance.—If two points be kept at a constant difference of potential, and be joined by a homogeneous conductor of uniform cross-section, it is found that the current in the conductor is directly proportional to its cross-section and inversely to its length. The current also depends upon the nature of the conductor. If conductors of similar dimensions, but of different materials, are used, the current in each is proportional to a quantity called the *specific conductivity* of the material. The numerical value of the current set up in a conducting cube, with edges of unit length, by unit difference of potential between two opposite faces, is the measure of the conductivity of the material of the cube. The reciprocal of this number is the *specific resistance* of the material. If ρ represent the specific resistance of the conducting material, S the cross-section, and l the length of a portion of the conductor of uniform cross-section between two points at potentials V_1 and V_2, Ohm's law for this special case is presented in the formula

$$I = \frac{S(V_1 - V_2)}{l\rho}. \tag{99}$$

The specific resistance is not perfectly constant for any one material, but varies with the temperature. In metals the specific resistance increases with rise in temperature; in liquids and in carbon it diminishes with rise in temperature. Upon this fact of change of resistance with temperature is based a very delicate instrument, called by Langley, its inventor, the *bolometer*, for the measurement of the intensity of radiant energy.

276. Joule's Law.—If we modify the equation $H = I(V_1 - V_2)$ by the help of Ohm's law, we obtain

$$H = I^2R. \tag{100}$$

The heat developed in a homogeneous portion of any circuit is equal to the square of the current in the circuit multiplied by the resistance of that portion. This relation was first experimentally proved by Joule in 1841, and is known as *Joule's law*. It holds true for any homogeneous circuit or for all parts of a circuit which are

homogeneous. The heat which is sometimes evolved by chemical action, or by the Peltier effect, occurs at non-homogeneous portions of the circuit.

277. Counter Electromotive Force in the Circuit.—In many cases the work done by the current does not appear wholly as heat developed in accordance with Joule's law.

Besides the production of heat throughout the circuit, work may be done during the passage of the current, in the decomposition of chemical compounds, in producing movements of magnetic bodies or in heating junctions of dissimilar substances.

Before discussing these cases separately we will connect them all by a general law, which will at the same time present the various methods by which currents can be maintained. They differ from the simple case in which the work done appears wholly as heat throughout the circuit, in that the work done appears partly as energy available to generate currents in the circuit. To show this we will use the method given by Helmholtz and by Thomson. The total energy expended in the circuit in the time t, which is such that, during it, the current is constant, is IEt. It appears partly as heat, which equals I^2Rt by Joule's law, and partly as other work, which experiment proves is in every case proportional to I, and can be set equal to IA, where A is a factor which varies with the particular work done. Then we have $IEt = I^2Rt + IA$, whence

$$I = \frac{E - \dfrac{A}{t}}{R}. \tag{101}$$

It is evident from the equation that $E - \dfrac{A}{t}$ is an electromotive force, and that the original electromotive force of the circuit has been modified by work having been done by the current. In other words, the performance of the work IA in the time t by the circuit has set up a *counter electromotive force* $\dfrac{A}{t}$. The separated constituents of the chemical compound, the moved magnet, the heated junction, are all sources of electromotive force which oppose that

of the original circuit. If then, in a circuit containing no impressed electromotive force, or in which $E = 0$, there be brought an arrangement of uncombined chemical substances which are capable of combination, or if in its presence a magnet be moved, or if a junction of two dissimilar parts of the circuit be heated, there will be set up an electromotive force $\frac{A}{t}$, and a current $I = \frac{A}{tR}$. Any of these methods may then be used as the means of generating a current. The first gives the ordinary battery currents of Volta, the second the induced currents discovered by Faraday, and the third the thermoelectric currents of Seebeck.

This demonstration fails when applied to the case of the induction of one current by another, in consequence of the changes produced in both by their mutual interactions. The correct demonstration in this case can only be reached by the aid of the dynamical equations of the electromagnetic field.

278. Poynting's Theorem.—On the view that the current consists of the disappearance of tubes of force in the conductor, the energy developed in the circuit enters it from the dielectric. By choosing a very simple case we may determine the rate at which this energy moves through the dielectric and into the conductor. We will suppose the current maintained in a very long straight cylindrical wire stretched between two parallel and very large planes, which are kept at the potentials V_1 and V_2. In such an arrangement the tubes of force are cylindrical, passing perpendicularly between the two plates and parallel with the conductor joining them; the electrical force in such a field is everywhere the same. Now consider a plane parallel with the plates, and describe in it a circle having any radius r with its centre at the centre of the wire. Let N represent the number of tubes of force which pass through unit area in this plane, and v the velocity with which these tubes of force pass through the circumference of the circle of radius r. Then the number of tubes which pass through this circumference in unit time is $2\pi rvN$, and since the current is supposed to

be steady, this number is the same whatever be the radius of the circle; it therefore expresses also the number of unit tubes which enter the wire in unit of time.

We have already seen that the energy carried into the conductor by Q unit tubes, when the difference of potential is maintained constant, is equal to $Q(V_1 - V_2)$, that is, is twice the energy associated with these tubes when at rest. It has also been shown that the energy in unit length of a unit tube at rest is $\frac{F}{2}$, and F therefore measures the energy carried into the conductor by unit length of each unit tube. In the case before us, the energy transferred in one second through a cylindrical surface of unit height and of radius r, concentric with the wire, is $2\pi rvNF$. Now, on the view of the current here taken, the number of unit tubes which disappear in one second is equal to the current strength, so that $2\pi rvN = I$. The energy introduced through the cylindrical surface is therefore FI. Since in this case the difference of potential equals the electrical force multiplied by the length of the wire, the energy introduced into the whole wire is $I(V_1 - V_2)$. The energy which passes through unit area of the cylindrical surface is $\frac{FI}{2\pi r}$. It may be shown that the magnetic force due to the current at the distance r is $P = \frac{2I}{r}$, and hence the energy which passes through unit area may also be represented by $\frac{FP}{4\pi}$.

The example here given is a special case of a general theorem due to Poynting. This theorem asserts that the energy expended in the current enters the conductor from the dielectric, passing at right angles to the lines of electrostatic force and the lines of magnetic force, and that the amount of energy which passes perpendicularly through unit area is proportional to the electrostatic force and to the magnetic force.

CHAPTER IV.

CHEMICAL RELATIONS OF THE CURRENT.

279. Electrolysis.—It has been already mentioned that, in certain cases, the existence of an electrical current in a circuit is accompanied by the decomposition into their constituents of chemical compounds forming part of the circuit. This process, called *electrolysis,* must now be considered more fully. It is one of those treated generally in § 277, in which work other than heating the circuit is done by the current. That work is done by the decomposition of a body the constituents of which, if left to themselves, tend to recombine, is evident from the fact that, if they be allowed to recombine, the combination is always attended with the evolution of heat or the appearance of some other form of energy. The amount of heat developed, or the energy gained, is, of course, the measure of the energy lost by combination or necessary to decomposition.

Those bodies which exhibit electrolysis are always such as have considerable freedom of motion among their molecules. Ordinarily, they are liquids or solids in solution or fused. The discharge through gases is also probably accompanied by electrolysis. Bodies which can be decomposed were called by Faraday, to whom the nomenclature of this subject is due, *electrolytes.* The current is usually introduced into the electrolyte by solid terminals called *electrodes.* The one at the higher potential is called the positive electrode or *anode;* the other, the negative electrode, or *cathode.* The two constituents into which the electrolyte is decomposed are

called *ions.* One of them appears at the anode and is called the *anion*, the other at the cathode and is called the *cation.*

For the sake of clearness we will describe some typical cases of electrolysis. The original observation of the evolution of gas when the current was passed through a drop of water, made by Nicholson and Carlisle, was soon modified by Carlisle in a way which is still generally in use. Two platinum electrodes are immersed in water slightly acidulated with sulphuric acid, and tubes are arranged above them so that the gases evolved can be collected separately. When the current is passing, bubbles of gas appear on the electrodes. When they are collected and examined, the gas which appears at the anode is found to be oxygen, and that which appears at the cathode to be hydrogen. The quantities evolved are in the proportion to form water. This appears to be a simple decomposition of water into its constituents, but it is probable that the acid in the water is first decomposed, and that the constituents of water are evolved by a secondary chemical reaction.

An experiment performed by Davy, by which he discovered the elements potassium and sodium, is a good example of simple electrolysis. He fused caustic potash in a platinum dish, which was made the anode, and immersed in the fused mass a platinum wire as cathode. Oxygen was then evolved at the anode, and the metal potassium was deposited on the cathode. This is the type of a large number of decompositions.

If, in a solution of zinc sulphate, a plate of copper be made the anode and a plate of zinc the cathode, there will be zinc deposited on the cathode and copper taken from the anode, so that, after the process has continued for a time, the solution will contain a quantity of cupric sulphate. This is a case similar to the electrolysis of acidulated water, in which the simple decomposition of the electrolyte is modified by secondary chemical reactions.

If two copper electrodes be immersed in a solution of cupric sulphate, copper will be removed from the anode and deposited on the cathode, without any important change occuring in the character or concentration of the electrolyte. This is an example of the

special case in which the secondary reactions in the electrolyte exactly balance the work done by the current in decomposition, so that on the whole no chemical work is done.

280. Faraday's Laws.—The researches of Faraday in electrolysis developed two laws, which are of great importance in the theory of chemistry as well as in electricity:

(1) The amount of an electrolyte decomposed is directly proportional to the quantity of electricity which passes through it; or, the rate at which a body is electrolyzed is proportional to the current strength.

(2) If the same current be passed through different electrolytes, the quantity of each ion evolved is proportional to its chemical equivalent. The chemical equivalent is the weight of the radical of the ion in terms of the weight of the atom of hydrogen, divided by its valency.

If we define an *electro-chemical equivalent* as the quantity of any ion which is evolved by unit current in unit time, then the two laws may be summed up by saying:

The number of electro-chemical equivalents evolved in a given time by the passage of any current through any electrolyte is equal to the number of units of electricity which pass through the electrolyte in the given time.

The electro-chemical equivalents of different ions are proportional to their chemical equivalents. Thus, if zinc sulphate, cupric sulphate, and argentic chloride be electrolyzed by the same current, zinc is deposited on the cathode in the first case, copper in the second, and silver in the third. The amounts by weight deposited are in proportion to the chemical equivalents, 32.6 parts of zinc, 31.7 parts of copper, and 108 parts of silver.

Faraday's laws may also be stated in another form, in which the word "ion" has a different meaning. The process of electrolysis consists in the separation of each molecule of the electrolyte into its constituent radicals. Each of these radicals is called an ion. If the valency of the radical be 1, the ion is called a univalent ion; if it be n, the ion is either called an n-valent ion or n-uni-

valent ions. To illustrate, we know that when hydrogen is evolved from hydrochloric acid, HCl, its ion is univalent. Now when it is evolved from water, H_2O, we may either consider the H_2 as a bivalent ion or as two univalent ions. Similarly we may consider the O as a bivalent ion or as two univalent ions, though it can never be actually broken up into two such ions. We may consider a molecule, then, as made up either of two n-valent ions or of $2n$ univalent ions. The weight of each of the n-valent ions may be measured in terms of the weight of the hydrogen atom taken as a unit, and is the molecular weight of the ion. This weight divided by the valency n is the weight of the univalent ion. It may be called the *ionic weight.*

Now the passage of a current through different electrolytes evolves their constituents in amounts proportional to their molecular weights divided by their valencies. It therefore evolves the ions in proportion to their ionic weights, or it evolves the same number of univalent ions in each electrolyte. Faraday's two laws may therefore be summed up in the statement that the number of univalent ions evolved by a current in any electrolyte is proportional to the quantity of current.

By this mode of considering electrolysis, we are led to the conclusion that each pair of univalent ions liberated during electrolysis is associated with a pair of charges numerically equal and of opposite sign. These charges are called *ionic charges.* An n-valent ion is associated with n ionic charges. If we use the conception of tubes of force, each positive univalent ion may be considered as the origin of a tube of force which terminates on a negative ion. Since the ionic charges are all equal, these tubes may be taken as unit tubes, which are no longer defined arbitrarily, but are based upon a constant of Nature.

281. The Voltameter.—These laws were used by Faraday to establish a method of measuring current by reference to an arbitrary standard. The method employs a vessel containing an electrolyte in which suitable electrodes are immersed, so arranged that the products of electrolysis, if gaseous, can be collected and meas-

ured, or, if solid, can be weighed. This arrangement is called a *voltameter*. If the current strength be desired, the current must be kept constant in the voltameter by suitable variation of the resistance in the circuit during the time in which electrolysis is going on.

Two forms of voltameter are in frequent use.

In the first form there is, on the whole, no chemical work done in the electrolytic process. The system consisting of two copper electrodes and cupric sulphate as the electrolyte is an example of such a voltameter. The weight of the copper deposited on the cathode measures the current.

The second form depends for its indications on the evolution of gas, the volume of which is measured. The water voltameter is a type, and is the form especially used. The gases evolved are either collected together, or the hydrogen alone is collected. The latter is preferable, because oxygen is more easily absorbed by water than hydrogen, and an error is thus introduced when the oxygen is measured.

282. Measure of the Counter Electromotive Force of Decomposition.—In the general formula developed in § 277 the quantity IA represents the energy expended in the circuit which does not appear as heat developed in accordance with Joule's law. In the present case it is the energy expended during electrolysis in decomposing chemical compounds and in doing mechanical work. In many cases the mechanical work done is not appreciable; but when a liquid like water is decomposed into its constituent gases, work is done by the expansion of the gases from their volume as water to their volume as gases. In many cases some of this energy is also used in keeping the temperature of the electrolyte constant. These cases occur when the electromotive force developed varies with the temperature.

In case no such variation with the temperature occurs, we may calculate the electromotive force developed in terms of heat. Let e represent the electro-chemical equivalent of one of the ions, and θ the heat evolved by the combination of a unit mass of this ion with

an equivalent mass of the other ion, in which is included the heat equivalent of the mechanical work done if the state of aggregation change. Then I will represent the number of electro-chemical equivalents evolved in unit time, and $Ie\theta t$ will represent the energy expended in the time t, which appears as chemical separation and mechanical work. This is equal to IA; whence $A = e\theta t$. All these quantities are measured in absolute units. The quantity $e\theta$ represents the energy required to separate the quantity e of the ion considered from the equivalent quantity of the other ion, and to bring both constituents to their normal condition. Now, $\frac{A}{t}$ represents the counter electromotive force set up in the circuit by electrolysis. Hence the electromotive force set up in the electrolytic process may be measured in terms of heat units.

It often is the case that the two ions which appear at the electrodes are not capable of direct recombination, as has been tacitly assumed in the definition of θ. A series of chemical exchanges is always possible, however, which will restore the ions as constituents of the electrolyte, and the total heat evolved for a unit mass of one ion during the process is the quantity θ.

The theory here presented is abundantly verified by the experiments of Joule, Favre and Silbermann, Wright, and others. The extension of the theory to cases in which the electromotive force varies with the temperature was made by Helmholtz.

283. Positive and Negative Ions.—Experiment shows that certain of the bodies which act as ions usually appear at the cathode, and certain others at the anode. The former are called *electro-positive* elements; the latter, *electro-negative* elements. Faraday divided all the ions into these two classes, and thought that every compound capable of electrolysis was made up of one electro-positive and one electro-negative ion. But the distinction is not absolute. Some ions are electro-positive in one combination and electro-negative in another. Berzelius made an attempt to arrange the ions in a series, such that any one ion should be electro-positive to all those above it and electro-negative to all those below it. There

is no reason to believe that such a rigorous arrangement of the ions can be made.

284. Grotthus's Theory of Electrolysis.—The foundation of all the present theories of electrolysis is found in the theory published by Grotthus in 1805. He considered the constituent ions of a molecule as oppositely electrified to an equal amount. When the current passes, owing to the electrical attractions of the electrodes, the molecules arrange themselves in lines with their similar ends in one direction, and then break up. The electro-negative ion of one molecule moves toward the positive electrode and meets the electro-positive ion of the neighboring molecule, with which it momentarily unites. At the ends of the line an electro-negative ion with its charge is freed at the anode, and an electro-positive ion with its charge is freed at the cathode. This process is repeated indefinitely so long as the current passes.

Faraday modified this view, in that he ascribed the arrangement of the molecules, and their disruption, to the stress in the medium which was the cardinal point in his electrical theories. Otherwise he held closely to Grotthus's theory. He showed that an electrical stress exists in the electrolyte by means of fine silk threads immersed in it. These arranged themselves along the lines of electrical stress.

Other phenomena, however, show that Grotthus's hypothesis can only be treated as a rough illustration of the main facts.

Joule showed that during electrolysis there is a development of heat at the electrodes, in certain cases, which is not accounted for by the elementary theory above given. It must depend upon a more complicated process of electrolysis than the one we have described.

The results of researches on the so-called *migration* of the ions are also at variance with Grotthus's theory. If the electrolysis of a copper salt, in a cell with a copper anode at the bottom, be examined, it will be found that the solution becomes more concentrated about the anode and more dilute about the cathode. These changes can be detected by the color of the parts of the solution, and substantiated by chemical analysis. If this result be explained

by Grotthus's theory, the explanation furnishes at the same time a numerical relation between the ions which have wandered to their respective regions in the electrolyte which is not in accord with experiment.

It is an objection against Grotthus's theory, and indeed against Thomson's method given in § 282 of connecting chemical affinity and electromotive force, that, on those theories, it would require an electromotive force in the circuit greater than $\frac{A}{t}$, the counter electromotive force in the electrolytic cell, to set up a current, and that the current would begin suddenly, with a finite value, after this electromotive force is reached. On the contrary, experiments show that the smallest electromotive force will set up a current in an electrolyte and even maintain one constantly, though the current strength may be extremely small.

285. The Dissociation Theory of Electrolysis.—The foundations of a more satisfactory theory of electrolysis were laid by Clausius, who proceeded from the view with which he had become familiar by his study of the kinetic theory of gases, that the molecules of all bodies are in constant motion. He assumed that the collisions of the molecules of the electrolyte occasionally caused a separation of some of the molecules into their constituent ions, and that the province of the electromotive force in the electrolyte was to direct the motion of these ions toward their respective electrodes. A considerable extension of Clausius's theory has been made by Arrhenius and developed by Ostwald and others, in which the leading idea is, that the molecules of an electrolyte in solution are always separated to a greater or less extent into their constituent ions. In many cases, and always in very dilute solutions, the separation, according to this view, is complete. This theory is called the *dissociation theory* of electrolysis. The ions, however, are not in the condition of the constituent parts of a molecule which have been dissociated at a high temperature (§ 219), but possess certain peculiar electrical and chemical properties. It has been proposed to call their condition in solution *ionization*. This term certainly possesses advantages, but

it has not yet come into common use, and we will therefore retain the term *dissociation.*

We have already seen that a current in an electrolyte may be considered as the transfer of charges on the moving ions. If the ions in solution be dissociated from each other, and if the effect of the electromotive force in the circuit be merely directive, it is plain that the quantity of current transferred will depend on the relative velocity with which the ions move past each other in the solution as well as on their number. Starting with this conception, we will show that the conductivity of an electrolyte is proportional to the sum of the velocities of its ions. The discovery of this fact by Kohlrausch laid the foundation for the dissociation theory.

Let us suppose a series of electrolytic cells, each one of which is a cubical box with sides of unit length, and so arranged that a current passes in them between two opposite faces which serve as electrodes. The column of the electrolyte between the electrodes is then one centimetre long and has a cross-section of one square centimetre. Let the electrolytes used in these cells be prepared by dissolving in equal volumes of the same solvent masses of the substances to be decomposed which are proportional to the sums of the ionic weights of their constituent ions (§ 280). Equal volumes of these solutions will then contain the same number of univalent ions.

If a current be sent through the series of cells containing these solutions, the same number of univalent ions will be liberated in each. The difference of potential between the terminals of the cells will be in general different for each of them. We have from Ohm's law the relation $I = k\,(V_1 - V_2)$, where the current I is the same for each cell and the difference of potential $V_1 - V_2$ and the conductivity k (§ 275) different for the different cells. Now consider a cross-section in one of the cells parallel with the electrodes; let u and v represent the velocities of the ions evolved in this cell. Let $2M$ represent the number of univalent ions in the cell, and let c represent the ionic charge. Now the relative velocity of the ions which pass through the cross-section taken in the cell is $u + v$; the

number of ions which pass through that cross-section in unit time in both directions is therefore $M(u+v)$ and the quantity of electricity carried through with them in both directions is $cM(u+v)$. But this quantity is equal to the current strength I, and therefore $cM(u+v) = k(V_1 - V_2)$; or $u + v = \frac{k(V_1 - V_2)}{cM}$. Now cM is the quantity of current required to decompose the molecules in the cell, or the mass which is in solution in unit volume of the electrolyte; it may therefore be directly determined. Since equal volumes of the electrolytes contain the same number of univalent ions, this quantity of current is the same for all the cells, and since, with a known value of I, we may determine the value of k in each case by observations of $V_1 - V_2$, the formula just obtained enables us to determine $u+v$.

This formula may be more conveniently used in another form. Let n represent the weight of the hydrogen evolved by unit current in unit time, and m the chemical equivalent of one of the products of electrolysis in the cell. Then mn represents the weight of that product evolved by unit current in unit time, and $\frac{1}{mn}$ represents the current that will evolve unit weight in unit time. Now the electrolytes are prepared so that the weights of the constituents in the cells are given by Nm, where N is a number which is the same for all the cells. The current that will evolve these weights in the respective cells is therefore equal to $\frac{N}{n}$, and this current has been shown to be equal to cM. Using this value of cM in the equation for $u+v$, we obtain $u + v = \frac{nk(V_1 - V_2)}{N}$. In the experiments of Kohlrausch the difference of potential $V_1 - V_2$ was the same for all the cells, and the value of $\frac{k}{N}$ determined for each cell. The values of $u+v$ could then be calculated. The ratio $\frac{k}{N}$ is called the *molecular conductivity*.

Now in order to determine the values of u and v, we need them combined in another relation; this relation may be obtained from a study of the migration of the ions. For, consider a row of molecules in the electrolyte stretching between the electrodes, of which the ions are moving independently, the positive ions to the right with the velocity u, and the negative ions to the left with the velocity v. Let n represent the number of ions of either sort in unit length of this line. At the end of the short time t the relative displacement of the rows of ions will be $(u + v)t$, and the number of ions freed at either end will be the same and equal to $n(u + v)t$. Though the number of ions which are freed at either end is the same, the loss of molecules or of pairs of associated ions is different at the two ends. If a line be drawn perpendicularly across the line of molecules, the number of ions which pass to the right, and therefore the number of molecules lost on the left of this line is nut, while the number of molecules lost on its right is nvt. If, therefore, we measure the diminution of the substance decomposed at each electrode, the ratio of the values found will be the ratio of the velocities u and v of the constituent ions. The ratio of one of these losses or diminutions to the sum of them both, or the ratio of the velocity of the corresponding ion to the sum of the velocities of the two ions, is called the *migration constant* of the ion. The migration constants have been determined for many ions by Hittorf, Nernst, and others. By combining the ratios of the velocities thus found with the sums of the velocities found by Kohlrausch, the velocities may be separately determined. It is found that the velocity of any one ion is the same, whatever be the electrolyte of which it forms a part, provided the solution be sufficiently dilute. This result is a strong confirmation of the theory of the independent motion of the ions upon which the calculations are based.

In many cases, especially when the solution is not very dilute, the molecular conductivity is found to be less than that assigned by theory on the assumption that all the ions of the electrolyte are dissociated. This discrepancy is explained by Arrhenius by the

assumption that in this case the dissociation is not complete; the ratio of the molecular conductivity found in such cases to the molecular conductivity at very great dilutions, in which case the dissociation is assumed to be complete, is taken as the measure of the dissociation in the solution. A similar theory of partial dissociation was assumed to account for the departures from the normal laws of osmotic pressure (§§ 94, 95), of the lowering of the freezing-point (§ 197), and of the lowering of vapor pressure (§ 204).

The agreement between the conclusions reached by these entirely independent methods with regard to the extent of dissociation is strong evidence in favor of the hypothesis upon which the calculations are based. Starting with the same hypothesis, other relations have been theoretically discovered among the physical properties of solutions which have been confirmed by experiment. The dissociation theory of solution and of electrolysis is not yet fully established, but it furnishes by far the most satisfactory explanation of the nature and behavior of solutions.

286. Voltaic Cells.—From the discussion given in § 277 it is obvious that, if an arrangement be made, in a circuit, of substances capable of uniting chemically and such as would result from electrolysis, there will result an electromotive force in such a sense as to oppose the current which would effect the electrolysis. If, then, the electrodes of an electrolytic cell in which this electromotive force exists be joined by a wire, a current will be set up through the wire in the opposite direction to the one which would continue the electrolysis, and the ions at the electrodes will recombine to form the electrolyte. There is thus formed an independent source of current, the *voltaic cell.* The electrode in connection with the electro-negative ion is called the *positive pole*, and that in connection with the electro-positive ion the *negative pole.*

Thus, if after the electrolysis of water in a voltameter, in which the gases are collected separately in tubes over platinum electrodes, the electrodes be joined by a wire, a current will be set up in it, and the gases will gradually, and at last totally, disappear, and the current will cease. The current which decomposes the water is

conventionally said to flow through the liquid from the anode to the cathode, from the electrode above which oxygen is collected to the electrode above which hydrogen is collected. The current existing during the recombination of the gases flows through the liquid from the hydrogen electrode to the oxygen electrode, or outside the liquid from the positive to the negative pole. Such an arrangement as is here described was devised by Grove, and is called the *Grove's gas battery.*

A combination known as *Smee's cell* consists of a plate of zinc and one of platinum, immersed in dilute sulphuric acid. It is such a cell as would be formed by the complete electrolysis of a solution of zinc sulphate, if the zinc plate were made the cathode. When the zinc and platinum plates are joined by a wire, a current is set up from the platinum to the zinc outside the liquid, and the zinc combines with the acid to form zinc sulphate. The hydrogen thus liberated appears at the platinum plate, where, since the oxygen which was the electro-negative ion of the hypothetical electrolysis by which the cell was formed does not exist there ready to combine with it, it collects in bubbles and passes up through the liquid. The presence of this hydrogen at once lowers the current from the cell, for it sets up a counter electromotive force, and also diminishes the surface of the platinum plate in contact with the liquid, and thus increases the resistance of the cell. It may be partially removed by mechanical movements of the plate or by roughening its surface. The counter electromotive force is called the *electromotive force of polarization.* It occurs soon after the circuit is joined up in all cells in which only a single liquid is used, and very much diminishes the currents which are at first produced.

Advantage is taken of secondary chemical reactions to avoid this electromotive force of polarization. The best example, and a cell which is of great practical value for its cheapness, durability, and constancy, is the *Daniell's cell.* Two liquids are used—solutions of cupric sulphate and zinc sulphate. They are best separated from one another by a porous porcelain diaphragm. A plate

of copper is immersed in the cupric sulphate, and a plate of zinc in the zinc sulphate. The copper is the positive pole, the zinc the negative pole. When the circuit is made, and the current passes, zinc is dissolved, the quantity of zinc sulphate increases and that of the cupric sulphate decreases, and copper is deposited on the copper plate. To prevent the destruction of the cell by the consumption of the cupric sulphate, crystals of the salt are placed in the solution. The electromotive force of this cell is evidently due to the loss of energy in the substitution of zinc for copper in the solution of cupric sulphate.

The *secondary cell of Planté* is an example of a cell made directly by electrolysis, as has been assumed in the preliminary discussion. The electrodes are both lead plates, and the electrolyte dilute sulphuric acid. When a current is passed through the cell, the oxygen evolved on the anode combines with the lead to form peroxide of lead, which coats the surface of the electrode. When the cell is inserted in a circuit, a current is set up, the peroxide is reduced to a lower oxide, and the metallic lead of the other plate is oxidized.

Cells of this sort, which have been constructed directly by coating lead plates with the proper oxides of lead, are called *storage cells.* They may be put in condition for use by sending a current through them in the proper direction. The sulphate of lead formed plays an important part in the operation of these cells.

The *Latimer-Clarke standard cell* is of great value as a standard of electromotive force. The positive pole consists of pure mercury, which is covered by a paste made by boiling mercurous sulphate in a saturated solution of zinc sulphate. The negative pole consists of pure zinc resting on the paste. Contact with the mercury is made by means of a platinum wire. As no gases are generated, this cell may be hermetically sealed against atmospheric influences. According to the measurements of Rayleigh, the electromotive force of this cell is very constantly $1{,}435 \cdot 10^8$ C. G. S. electromagnetic units at 15° Cent.

287. Theories of the Electromotive Force of the Voltaic Cell.—The plan followed in the preceding discussions has rendered it unnecessary for us to adopt any theory to explain the cause of the electromotive force of the voltaic cell. The different theories which have been advanced may be classed under one of two general theories, the contact theory and the chemical theory. On the *contact theory*, as advanced by Volta and supported by Thomson and others, the difference of potential which exists between two heterogeneous substances in contact is due to molecular interactions across the surface of contact, or, as it is commonly stated, is due merely to the contact. The *chemical theory*, as advocated by Faraday and Schönbein, holds that the difference of potential considered cannot arise unless chemical action or a tendency to chemical action exist at the surface of contact.

Numerous experiments have shown that the sum of all the differences of potential at the surfaces of contact of the various substances making up any voltaic cell is equal to the electromotive force of that cell. This is true even when the cell is formed solely of liquid elements. On the contact theory, this electromotive force is due merely to the several contacts, while the chemical actions of the cell begin only when the circuit is made, and supply the energy for the maintenance of the current. On the chemical theory the chemical action of the medium is concerned in the production of the difference of potential observed.

On either theory it is clear that the energy maintaining the current must have its origin in the chemical actions which go on in the voltaic cell.

288. The Electrical Double-sheet.—Suppose two plates of different materials, say one zinc and the other copper, joined by a wire and placed opposite each other like the plates of a condenser: as stated in the last section, a difference of potential then exists between them. The charge on one of them is given by $\frac{S(V_1 - V_2)}{4\pi d}$ (§ 259, (Eq. 92)). The difference of potential will remain the same, whatever be the distance between the plates, so that the charges on

the plates and the distance between them vary inversely. When the faces of the two plates are in contact, that is, are separated by molecular distances, these charges become very great. Such an arrangement of equal and opposite charges, distributed over the surfaces of two bodies in contact and separated by a distance comparable with the distance between the molecules, was called by Helmholtz an *electrical double-sheet*. It evidently presents some analogies to the magnetic shell.

The charges making up the double-sheet cannot be detected by separating a plate of zinc from a plate of copper with which it has been in contact and examining the separate plates, because the separation cannot be effected so uniformly that no discharge takes place between the two bodies. If, however, those faces of the zinc and copper plates which are contiguous be insulated from each other by a thin layer of shellac and contact made between the plates by means of a metallic wire, so that a difference of potential is set up between them, on removal of the wire and separation of the plates they are found to possess charges of considerable magnitude.

We may explain in this way electrification by friction. We may assume that the two bodies rubbed together acquire different potentials by contact; the friction forces large areas of their surfaces into close proximity, and the charges upon those surfaces become very great; because the bodies ordinarily used for producing electrification by friction are nonconductors, the charges on their surfaces are not recombined as the bodies are separated, so that each body retains a large free charge.

A similar electrical double-sheet will exist on the surfaces of contact between a liquid and a metal. An arrangement by which the effects due to this double-sheet may be observed was invented by Lippmann. It consists of a vertical glass tube drawn out at its lower end in a capillary tube. The capillary tube dips into dilute sulphuric acid, which rests on mercury in the bottom of the vessel containing it. Mercury is poured into the vertical tube until its pressure is such that the capillary portion of the tube is nearly filled with it. When the mercury in the vessel is joined with the

positive pole of a voltaic cell, and that in the tube with the negative pole, the meniscus in the capillary tube moves upward, in the sense in which it would move if its surface tension were increased.

This movement may be explained as follows: An electrical double-sheet will be formed on the curved surface of contact of the mercury and acid in the capillary tube, and the interaction of the parts of this double-sheet will give rise to an electrical pressure (§ 256), that diminishes the apparent surface tension in that surface. If a weak current be sent through the solution, the difference of potential between the liquid and mercury will be diminished or increased by the ionic charges transferred by the current, according as the current flows in one direction or the other. The apparent surface tension will be altered and the end of the mercury column will be displaced; the true surface tension of the surface will be efficient only when the mercury and solution are at the same potential, and this surface tension will be a maximum. The experiments of Helmholtz and A. König have shown that such a maximum exists in a way consistent with this view.

The arrangement described can manifestly be used to produce the effects just discussed only when the electromotive force introduced into the circuit is less than that required to cause active decomposition of the electrolyte.

Lippmann constructed an apparatus similar to the one described, with the addition of an arrangement by which pressure can be applied to force the end of the mercury column in the capillary tube back to the fixed position which it occupies when no electromotive force is introduced into the circuit. He found that when small electromotive forces were introduced, the pressures required to bring the end of the column back to the fixed position were proportional to the electromotive forces. He hence called this apparatus a *capillary electrometer*.

Lippmann also found that if the area of the surface of separation between the mercury and the liquid in the capillary tube were altered by increasing the pressure and driving the mercury down the tube, a current was set up in a galvanometer inserted in the

circuit, in a sense opposite to that which would change the area of the meniscus back to its original amount.

The electrical double-sheet produced by contact of a liquid and a solid serves also to explain the phenomenon of *electrical endosmose.*

It is found that, if an electrolyte be divided into two portions by a porous diaphragm, there is a transfer of the electrolyte toward the cathode, so that it stands at a higher level on the side of the diaphragm nearer the cathode than on the other. This fact was discovered by Reuss in 1807, and has been investigated by Wiedemann and Quincke. They found that the amount of the electrolyte transferred is proportional to the current strength, and independent of the extent of surface or the thickness of the diaphragm. Quincke has also demonstrated a flow of the electrolyte toward the cathode in a narrow tube, without the intervention of a diaphragm. Those electrolytes which are the poorest conductors show the phenomenon the best. In a very few cases the motion is towards the anode. The material of which the tube is composed influences the direction of flow. It has also been shown that solid particles move in the electrolyte, usually towards the anode.

Helmholtz showed that these movements can be explained by taking into account the interaction between the ionic charges and the double-sheet, and the viscosity of the liquid.

CHAPTER V.

THE MAGNETIC RELATIONS OF THE CURRENT.

289. The Magnetic Field of a Current.—Soon after the discovery by Oersted of the force exerted by an electrical current on a magnet, Biot and Savart instituted experiments to discover the law of this force. They suspended a small magnet near a long vertical wire through which a current was passing and counteracted the earth's magnetic field by magnets, so that, if no current passed through the wire, the small magnet was free from any directive force. When the current passed, the magnet placed itself at right angles to the plane containing the wire and the centre of the magnet. By oscillating the magnet, the strength of the magnetic field acting upon the magnet was found to be directly proportional to the strength of the current and inversely proportional to the distance between the magnet and the current.

It follows at once, from the position assumed by the short magnet, that the lines of magnetic force set up by the current are circles with their centres in the current. The relation between the direction of the current and the direction of the lines of force set up by it may be described in several ways. Ampère's rule is as follows: If the observer imagine himself swimming with the current and looking toward the magnet, the north pole of the magnet tends to move toward his left. Maxwell's rule is, that the direction of the current and the direction of its lines of force are related as the directions of translation and rotation of a right-handed screw. This rule is the one now commonly used. By supposing the wire carrying the current bent around into a closed curve, it will easily

be seen that the relation between the current and the lines of force is also that between the lines of force and the current; that is, the direction of the lines of force and the direction of the current are related as the directions of translation and rotation of a right-handed screw. A simple rule equivalent to these others is as follows: Let the conductor carrying the current be grasped with the right hand and the thumb extended along it in the direction of the current; the fingers then point in the direction of the lines of force.

In accordance with the views prevalent at the time, Biot supposed that the action of the current upon a magnet pole was due to the independent action of each element of the current. He showed that the results of his experiments were consistent with the assumption that a force acts between a magnet pole m and an element ds of the current i, at the distance r from the magnet pole and making an angle α with r, equal to $\frac{mi \sin \alpha \, ds}{r^2}$. At present we no longer consider the current as acting at a distance in accordance with this formula, but consider it rather as setting up a magnetic field, and we express its action upon a magnet pole in terms of the field which it sets up. We will return to the consideration of Biot's formula after developing this method.

It was shown by Ampère, and later by Weber, that a very small closed plane circuit sets up a magnetic field similar to that about a small magnet placed with its centre at the centre of the circuit, with its axis normal to the plane of the circuit. This magnet may be replaced by a magnetic shell with its edge coincident with the circuit, without altering the magnetic field. At all points outside the shell its magnetic field is similar to the magnetic field set up by the current; at those points in the field occupied by the substance of the shell the conditions are not the same in both cases. The potential of a shell at a point outside it is $j\omega$ (§ 243), where j is the strength of the shell and ω is the solid angle subtended by the shell. This is also the potential of the current, if the current be measured in such units that the current strength $i = j$. Now a shell of finite area may be built up of a number of elementary

shells, and likewise a current in a circuit coincident with the boundary of the finite shell may be built up of the elementary circuits corresponding to the elementary shells; for the current in each of the elementary circuits will be everywhere neutralized by the equal and opposite currents of the contiguous circuits except at the boundary of the surface occupied by the circuits. At the boundary the currents of the elementary circuits are in the same direction, and are not neutralized by other currents; they are therefore equivalent to the current in the circuit coincident with the boundary of the shell. This reasoning is plain from Fig. 84. If the strength of a finite shell be constant, the potential of the shell is $j\Omega$, where Ω is the solid angle subtended by the shell from an external point. The potential of the equivalent current is therefore $i\,\Omega$.

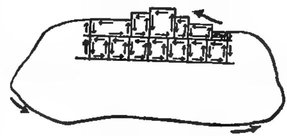

FIG. 84.

290. Multiply-valued Potential of the Current.—There is an important difference between the potential due to a current and that due to its equivalent magnetic shell, owing to the fact that the substance of the shell interrupts the field so that the potential within it does not follow the same law as that outside it. If we suppose that the shell is plane, the potential at a point on its positive face is $2\pi j$ and that at the corresponding point on the negative face is $-2\pi j$, so that the work done in transferring a unit pole from one point to the other is $4\pi j$ (§ 243). The pole can only be brought back to the point from which it started by carrying it around the edge of the shell and by doing upon it the work $4\pi j$, so that when it is returned to the starting-point the work done upon it is zero. If, on the other hand, the pole is moving under the influence of the circuit equivalent to the magnetic shell, the work done in transferring it outside the circuit from the positive to the negative face is equal to $4\pi i$. But it is not necessary to carry it again outside the circuit to return it to the starting-point. This may be accomplished by an infinitesimal displacement through the plane of the circuit; and since the force is everywhere finite, no

work is done in this displacement. The system then returns to its original condition, and work equal to $4\pi i$ is done upon the pole. This is expressed by saying that the potential of a closed current is *multiply-valued.* The work done during any movement depends not only on the position of the initial and final points in the path, as in the case of the ordinary single-valued gravitational, electrical, and magnetic potentials, but also on the path traversed by the moving magnet pole. Every time the path encloses the current, work equal to $4\pi i$ is done. The work done in moving by a path which does not enclose the current, from a point where the solid angle subtended by the circuit is Ω' to one where it is Ω, is, as in the case of the magnetic shell, equal to $i(\Omega' - \Omega)$. If the path further enclose the current n times, the work done is $4\pi n i$, so that the total work done, or the total difference of potential between the two points, is

$$V' - V = i(\Omega' - \Omega + 4\pi n), \tag{102}$$

where n may have any value from 0 to infinity.

The fact that the potential of a current is multiply-valued is well illustrated by any one of a series of experiments due to Faraday. If we imagine a wire frame forming three sides of a rectangle to be mounted on a support so as to turn freely about one of its sides as a vertical axis, while the free end of the opposite side dips in mercury contained in a circular trough of which the axis of rotation passes through the centre, and if we suppose a current to be sent through the axis and the frame, passing out through the mercury; then if a magnet be placed vertically with its centre on the level of the trough, and with either pole confronting the frame, the frame will rotate continuously about the axis.

Other arrangements are made by which more complicated rotations of circuits can be effected. If the circuit be fixed and the magnet movable, similar arrangements will give rise to motions of the magnet or to rotations about its own axis.

291. Electromagnetic Unit of Current.—The relation which has been discussed between a circuit and the equivalent magnetic shell affords a means of defining a unit of current different from that

before defined in the electrostatic system. That current is defined as the *unit current*, which will set up the same magnetic field as that due to a magnetic shell of which the edge coincides with the circuit, and the strength is unity.

The unit based upon these definitions is called the *electromagnetic* unit of current. It is fundamental in the construction of the electromagnetic system of units, in just the same way as the unit of quantity is fundamental in the electrostatic system.

The dimensions of current in the electromagnetic system are the same as those of strength of shell, that is, $[i] = M^{\frac{1}{2}}L^{\frac{1}{2}}T^{-1}$.

In practice, another unit of current is used, called the *ampere*. It contains 10^{-1} absolute electromagnetic units.

292. Energy of a Current in a Magnetic Field.—It has been shown (§ 289) that the potential at a point in a field due to a current is equal to $i\Omega$, where Ω is the solid angle subtended by the circuit at the point. If a pole of strength m be placed at that point, the potential energy of the pole is equal to $mi\Omega$; and since the same amount of work will be done if the circuit be fixed and the pole moved to an infinite distance as is done if the pole be fixed and the circuit removed to an infinite distance, the expression $mi\Omega$ also measures the energy of the circuit in the field due to the pole. Now $4\pi m$ is the number of unit tubes which proceed from the pole, and therefore $m\Omega$ is the number of unit tubes which pass through the circuit. We may therefore express the energy of the circuit due to the pole by iN, where N is the number of unit tubes which pass through the circuit. If the circuit be placed in any magnetic field, the forces in the field and the tubes of induction may be considered as due to an assemblage of magnetic poles, to each of which the proposition just stated applies. The energy of the circuit in any magnetic field is therefore given by iN, where N is the number of tubes of induction due to the field which pass through the circuit. N is positive when the tubes of induction of the field pass through the circuit in a direction opposed to that of the tubes of induction of the circuit; and is negative when they pass through in the same direction as that of the tubes of the

circuit. These statements hold true not only in the case supposed, in which the field is homogeneous, but also when the field contains masses of magnetizable matter which distort the tubes of induction.

293. Energy of a Current in its own Field.—When the circuit is traversed by a current, a magnetic field is present around it, and the circuit possesses energy in consequence of the presence of that field; we may calculate an expression for this energy, if we assume that it is distributed in the tubes of induction around the circuit according to the law developed in § 248. Let the number of unit tubes of induction which pass through the circuit be represented by N. The unit of length of each of these tubes contains an amount of energy equal to $\frac{R}{8\pi}$, where R is the resultant magnetic force. Any one tube therefore contains energy equal to $\frac{\Sigma R \Delta l}{8\pi}$, where Δl is an element of length of the tube and the summation is extended over the whole tube. But $\Sigma R \Delta l$ equals the work done in carrying a unit pole over the whole length of the tube. The tube is a continuous closed tube enclosing the circuit, and the work done in carrying a unit pole over such a closed curve enclosing the circuit is equal to $4\pi i$ (§ 290), so that $\Sigma R \Delta l = 4\pi i$. The energy of each unit tube is therefore equal to $\frac{1}{2}i$, and the energy of all the tubes belonging to the circuit is equal to $\frac{1}{2}iN$. Therefore, the energy of the circuit, due to its own current, is equal to one half the product of the current and the number of tubes of induction which pass through the circuit. Now we know by experiment that the magnetic force due to a current or the number of tubes of induction which pass through unit area is proportional to the current. Let L represent the number of tubes of induction which pass through the circuit when it is traversed by unit current; L is called the *coefficient of self-induction*. Then $N = Li$, and the energy of the circuit equals $\frac{1}{2}Li^2$.

294. Energy of Two Circuits.—If two circuits be present in a field, each of them possesses a certain amount of energy due to the magnetic field set up by the other. Let N_1 represent the number

of tubes of induction which pass through circuit 1 in consequence of the current in the other circuit. Then the energy of circuit 1, in consequence of the presence of the other circuit, is i_1N_1; the energy of circuit 2, in consequence of the presence of circuit 1, will be similarly i_2N_2. Now since there will be no mutual action between the circuits if either one of them is removed to an infinite distance, the work done in removing one of them is equal to its energy due to the presence of the other; and since, manifestly, the same amount of work is done if either one of the circuits be kept fixed and the other removed to an infinite distance, their energies must be equal, or $i_1N_1 = i_2N_2$. Now N_1 and N_2 are proportional respectively to the currents in circuits 2 and 1. Let M_1 represent the number of tubes of induction which pass through circuit 1 in consequence of unit current in circuit 2, and M_2 the corresponding number which pass through circuit 2 in consequence of unit current in circuit 1. Then $i_1i_2M_1 = i_1i_2M_2$ or $M_1 = M_2 = M$. The number of tubes of induction which pass through either circuit in consequence of a unit current in the other circuit is the same; the coefficient M which expresses this number is called the *coefficient of mutual induction.*

The energy of two circuits is equal to the energy which they possess due to their own currents, and the energy which each of them possesses due to the current in the other. If L and N are their coefficients of self-induction and M their coefficient of mutual induction, their energy is equal to $\frac{1}{2}Li_1^2+Mi_1i_2 + \frac{1}{2}Ni_2^2$. This energy may be represented as divided between the two circuits by the equivalent formula $\frac{1}{2}i_1(Li_1 + Mi_2)+ \frac{1}{2}i_2(Mi_1+Ni_2)$, where the terms represent one half the current in the circuit multiplied by the number of tubes of induction which pass through the circuit.

295. Motion of a Circuit in a Magnetic Field.—The motion of a circuit in a magnetic field, if the current in it be supposed constant, may always be found, from the results of the preceding sections, by the help of the general rule that the motion is such as to make the energy of the circuit as small as possible. In the simple case where the magnetic field is due to a north magnet pole, the

potential of the pole is positive when the current, as seen from the pole, is directed counterclockwise, that is, when the face of the circuit which confronts the pole is its *positive face*, which corresponds to the north face of a magnetic shell. The energy of the circuit is also positive in this position. It becomes zero when the pole is brought into the plane of the circuit outside of it, and negative when the pole confronts the *negative face*. The energy of the circuit is therefore diminished by turning its negative face toward the pole and moving it up toward the pole. The tubes of induction of the pole then pass through the circuit in the *positive direction*, that is, in the same direction as the tubes of the circuit; and the motion is such as to include as many of the tubes of the pole in the circuit as possible. The rule thus illustrated is a general one: a circuit in a magnetic field tends to move so that as many of the tubes of the field as possible pass through it in the positive direction; or, more fully, it tends to move so that the difference between the tubes which pass through it in the positive direction and in the negative direction is as great as possible. In terms of the symbols already used, the motion is such as to make N negative and numerically as great as possible. From this rule it is easy to see that a circuit will be in stable equilibrium with a soft iron bar when the axis of the bar is normal to the circuit, the tubes of induction in the bar are in the same direction as those of the circuit, and the bar is as near the edge of the circuit as possible.

When the field is due to the presence of another circuit the motion is such as to set their tubes of induction in the same direction, and to include in each circuit as many of the tubes of the other as possible; that is, to make M negative and numerically as great as possible. When the circuits are thus placed, their currents are travelling in the same sense. Their mutual action may therefore be expressed by saying that currents travelling in the same sense attract, and in opposite senses repel, each other.

The action on a circuit in a field due to magnets, or the mutual action of two circuits, may be described in terms of the actions that would be exerted on the magnetic shells which are equivalent to

them. Or, we may consider the tubes of induction as tending to diminish in length and to repel each other, and describe the action on a circuit in terms of the tensions due to the tubes of induction. These various modes of description necessarily yield similar results.

296. Action of a Current on a Magnet Pole.—We will now show that the force between a magnet pole and a circuit carrying a current may be considered as the resultant of forces which act between the pole and the elements of the circuit, and that this action follows the law deduced by Biot directly from his experiments (§ 289). On the view we have taken, this representation of the action is an artificial one, the real action being due to the magnetic field associated with the circuit.

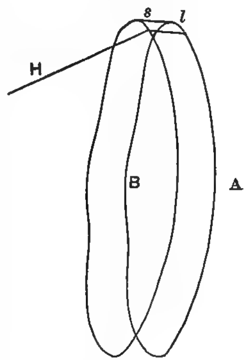

Fig. 85.

Let AB (Fig. 85) represent a circuit carrying the current ι, placed in a magnetic field in which the permeability is unity; let l represent the length of an element of the circuit, and H the strength of the magnetic field near that element. If N represent the number of unit tubes of force which pass through the circuit, the energy of the circuit is expressed by iN. Suppose the circuit displaced so that all parts of it move through the same small distance s. The number of tubes of force which pass through it after its displacement is represented by N'; the energy lost by the displacement is $\iota(N - N')$. This energy is equal to the work done upon the circuit by the forces of the field.

If we consider the closed surface bounded by the planes of the circuit in its two positions and by the cylinder traced by the circuit during its displacement, and remember that there is no free magnetism within this surface, the flux of force over it is zero (§ 56). And since the change in the number of unit tubes passing through the circuit measures the change in the flux of force through the circuit, it is evident that the change in the flux of force through

the circuit, or $N - N'$, is equal to the flux of force through the cylindrical surface. Now let θ represent the angle between s and l; the area traversed by l during its displacement is then $sl \sin\theta$. Let ϕ represent the angle between the normal to this area and the direction of the magnetic force H acting through it. Then $H \cos\phi$ is the component of the magnetic force normal to this area, and $Hsl \sin\theta \cos\phi$ is the flux of force through this area. The flux of force through the cylindrical surface is therefore given by $\Sigma Hsl \sin\theta \cos\phi = N - N'$. The energy lost by the displacement, or $i(N - N')$, is equal to $i\Sigma Hsl \sin\theta \cos\phi$; and since all parts of the circuit are displaced through the same distance s, this loss of energy is equal to the work which would be done on the circuit by a force acting in the direction of s and equal to $i\Sigma Hl \sin\theta \cos\phi$, or by a force acting on each element of the circuit equal to $iHl \sin\theta \cos\phi$. We may therefore consider the action of the magnetic field on the circuit as the resultant of an action of the magnetic field on each element of the circuit.

The magnitude and direction of the resultant force which acts on each element may be found as follows: The force $iHl \sin\theta \cos\phi$ is equal to zero when $\sin\theta = 0$, or when s and l coincide with each other; it is also equal to zero when $\cos\phi = 0$, or when the direction of H lies in the surface described by l. The resultant or maximum force which acts on the element is therefore at right angles to l and to H; the element l is urged to move at right angles to itself and to the magnetic force. The magnitude of the force acting on an element is obtained by supposing the element displaced in this direction, that is, along the normal to l and H. In this case we have $\sin\theta = 0$, and $\cos\phi = \sin\alpha$, where α is the angle between the element l and the direction of the magnetic force H. Substituting these values, the resultant force on the element is found to be equal to $iH \sin\alpha$.

In the special case in which the magnetic field is due to a single magnet pole of strength m, we have $H = \frac{m}{r^2}$, where r is the distance from the pole to the element of the circuit. The force

exerted by a magnet pole on an element of the circuit is therefore $\frac{mil}{r^2}\sin\alpha$, and this force urges the element to move at right angles to itself and to the line joining it with the magnet pole. Since the action between the pole and the circuit is mutual and the work done dependent only on their relative displacements, the force which each element of the circuit exerts on the pole is also equal to $\frac{mil}{r^2}\sin\alpha$, and tends to urge the pole to move at right angles to the plane containing it and the element of the circuit. This action on the magnet pole is the same as that deduced by Biot from his study of the force between a pole and a long straight current.

We will apply this theorem to determine the force due to a circular current on a magnet pole placed at a point on the line drawn normal to the plane of the circuit through its centre. The force on the circuit, and therefore the force on the pole, has been shown to be equal to $i\Sigma Hl\sin\theta\cos\phi$. In the case before us $H = \frac{m}{R^2}$, where m is the strength of the magnet pole and R the distance from the pole to the circuit. Since the elements of the circuit are symmetrical with respect to the pole, the force on the pole is along the line joining it to the centre of the circuit. The angle θ therefore equals $\frac{\pi}{2}$ and $\sin\theta = 1$; the angle ϕ is the angle between the radius of the circle and R; and $\cos\phi = \frac{r}{R}$, where r is the radius of the circle. The sum of all the elements of the circuit equals the circumference of the circle, or $2\pi r$. The force on the pole is therefore equal to $\frac{2\pi mir^2}{R^3}$.

If the magnet pole be placed at the centre of the circle, so that $R = r$, the force on it becomes $\frac{2\pi mi}{r}$. Let the radius of the circle be the unit length, or one centimetre; the force acting on the magnet pole is then $2\pi mi$, and if the magnet pole be the unit pole,

the force is $2\pi i$. If therefore the force exerted be equal to 2π, i will be equal to unity. We have thus arrived at another definition of unit current from the point of view of Biot's law. The unit current is defined to be that current which, flowing in a circle of unit radius, will exert upon a unit magnet pole at its centre a force equal to 2π dynes.

297. Ampère's Law for the Mutual Action of Currents.—The mutual action of two currents may also be considered as arising from forces between the elements of the currents. It was from this point of view that the action of currents was first investigated by Ampère. While the results obtained by him were not a unique solution of the problem, and must be regarded only as an artificial representation of the action between currents, they are yet of great interest. Without attempting to deduce Ampère's law, we will briefly consider the experiments upon which his deductions were based.

Ampère's method consists in submitting a movable circuit or part of a circuit carrying a current to the action of a fixed circuit, and in so disposing the parts of the fixed circuit that the forces arising from different parts exactly annul one another, so that the movable circuit does not move when the current in the fixed circuit is made or broken. In the first two of his experiments the movable circuit consists of a wire frame of the form shown in Fig. 86.

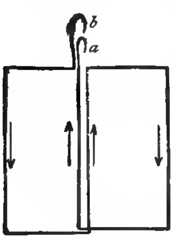

FIG. 86.

The current passes into the frame by the points a and b, upon which the frame is supported. It is evident that the two halves of the frame tend to face in opposite directions in the earth's magnetic field, so that there is no tendency of the frame as a whole to face in any one direction rather than any other. If a long straight wire be placed near to one of the extreme vertical sides of the frame and a current be sent through it, that side will move towards the wire if the currents in it and in the wire be in the same direction, and will move away from the wire if the currents be in opposite directions.

If now this wire be doubled on itself, so that near the frame there are two equal currents occupying practically the same position, but in opposite directions, then no motion of the frame can be observed when a current is set up in the wire. This is Ampère's *first case* of equilibrium. It shows that the forces due to two currents, identical in strength and in position, but opposite in direction, are equal and opposite.

If the portion of the wire which is doubled back be not left straight, but bent into any sinuosities, provided these be small compared with the distance between the wire and the frame, still no motion of the frame occurs when a current is set up in the wire. This is Ampère's *second case* of equilibrium. It shows that the action of the elements of the curved conductor is the same as that of their projections on the straight conductor.

To obtain the *third case* of equilibrium, a wire, bent in the arc of a circle, is arranged so that it may turn freely about a vertical axis passing through the centre of the circle of which the wire forms an arc, and normal to the plane of that circle. The wire is then free to move only in the circumference of that circle, or in the direction of its own length. Two vessels filled with mercury, so that the mercury stands above the level of their sides, are brought under the wire arc, and raised until conducting contact is made between the wire and the mercury in both vessels. A current is then passed through the movable wire through the mercury. Then if any closed circuit whatever, or any magnet, be brought near the wire, it is found that the wire remains stationary. The deduction from this observation is that no closed circuit tends to displace an element of current in the direction of its length.

In the fourth experiment three circuits are used, which we may call respectively A, B, and C. They are alike in form, and the dimensions of B are mean proportionals to the corresponding dimensions of A and C. B is suspended so as to be free to move, and A and C are placed on opposite sides of B, so that the ratio of their distances from B is the same as the ratio of the dimensions of A to those of B. If then the same current be sent through A and C,

and any current whatever through B, it is found that B does not move. The opposing forces due to the actions of A and C upon B are in equilibrium. From this *fourth case* of equilibrium is deduced the law that the force between two current elements is inversely as the square of the distance between them.

Ampère made the assumption that the action between two current elements is in the line joining them. From the four cases of equilibrium he then deduced an expression for the attraction between two current elements. It is

$$\frac{ii'\,ds\,ds'}{r^2}(2\cos\epsilon - 3\cos\theta\cos\theta'). \tag{103}$$

In this formula ds and ds' represent the elements of the two circuits, i and i' the strength of current in those circuits measured in electromagnetic units, r the distance between the current elements, ϵ the angles made by the two elements with one another, θ and θ' the angles made by ds and ds' with r or r produced, the direction of the two elements being taken in the sense of their respective currents.

298. Solenoids and Electromagnets.—Ampère also showed that the action between two small plane circuits is the same as that between two small magnetic shells, and that a circuit, or system of circuits, may be constructed which is the complete equivalent of any magnet. A long bar magnet may be looked on as made up of a great number of equal and similar magnetic shells arranged perpendicularly to the axis of the magnet, with their similar faces all in one direction. In order to produce the equivalent of this arrangement with the circuit, a long insulated wire is wound into a close spiral, straight and of uniform cross-section. The end of the wire is passed back through the spiral. When the current passes, the action of each turn of the spiral may be resolved into two parts—that due to the projection of the spiral on the plane normal to the axis, and that due to its projection on the axis. This latter component, for every turn, is neutralized by the current in the returning wire, and the action of the spiral is reduced to that of a number of similar plane circuits perpendicular to its axis.

Such an arrangement is called a *solenoid*. The poles of a solenoid of very small cross-section are situated at its ends, and it is equivalent to a bar magnet uniformly magnetized.

If a bar of soft iron be introduced into the magnetic field within a solenoid it will become magnetized by induction. This combination is called an *electromagnet*. Since the strength of the magnetic field varies with the strength of the current in the solenoid, and with the number of layers of wire wrapped around the iron core, the magnetization of any bar of iron whatever may be raised to its maximum by increasing the current and the number of turns of wire.

299. Ampère's Theory of Magnetism.—Ampère based upon these facts a theory of magnetism which bears his name. He assumed that around every molecule of iron there circulates an electrical current, and that to such molecular currents are due all magnetic phenomena. He made no hypothesis with regard to the origin or the permanency of these currents. The theory agrees with Weber's hypothesis that magnetization consists in an arrangement of magnetic molecules.

Ampère's theory admits of an explanation of diamagnetism, which was given by Weber. He assumes that all diamagnetic molecules are capable of carrying molecular currents, but that those currents, under ordinary conditions, do not exist in them. When, however, a diamagnetic body is moved up to a magnet an induced current due to the motion (§ 306) is set up in each molecule, and in such a direction that the molecules become elementary magnets, with their poles so directed towards the magnet in the field that there is repulsion between them. If this theory be true, it ought to be possible, as suggested by Maxwell, to lessen the intensity of magnetization of a body magnetized by induction, by increasing the strength of the field beyond a certain limit. No such effect has as yet been observed.

We may state the facts of magnetism in a way which is more in accordance with our view that the current is the result of actions in the medium by saying that each magnetic molecule is the origin of a certain definite number of tubes of induction. The existence

of these tubes is supposed to be connected with the peculiar motions which characterize the molecule of the magnetic body. Diamagnetism would then be explained by supposing a similar motion enforced upon the molecules of the other bodies in the field to an extent in each which depends upon the nature of the body.

300. The Hall Effect.—Hitherto it has been assumed that when currents interact, it is their conductors alone which are affected, and that the currents in the conductors are not in any way altered. Hall has, however, discovered a fact which seems to show that currents may be displaced in their conductors. If the two poles of a voltaic battery be joined to two opposite arms of a cross of gold-foil mounted on a glass plate, and if a galvanometer be joined to the other two arms at such points that no current flows through it, and if a magnet pole be brought opposite the face of the cross, a permanent current will be indicated by the galvanometer. The same effect appears in the case of other metals. The direction of the permanent current and its amount differ under the same circumstances for different metals. The coefficient which represents the amount of the Hall effect in any metal is called the *rotational coefficient* of that metal.

Since the rotational coefficients of such metals as have been tested agree in sign and in relative magnitude with their thermoelectric powers (§ 316), it is argued by Bidwell, v. Ettingshausen, and others that the Hall effect is due to thermoelectric action.

301. Currents in a Magnetic Field Due to Inequalities of Temperature.—If a thin strip of bismuth be placed in a magnetic field so that the magnetic force is normal to its surface, and if one of the edges of the strip be kept at a higher temperature than the other and the two ends of the strip joined by a wire in which a galvanometer is inserted, a continuous current will flow through the circuit. The direction of this current changes when the direction of the flow of heat changes or when the magnetic field is reversed. The strength of the current is different in different metals. These facts were discovered by v. Ettingshausen. Conversely, if a current be sent through the strip of bismuth placed in the magnetic field, there

will be a flow of heat across the strip and the temperatures of its edges will differ. This difference of temperature is due to a cooling of one edge of the strip. These effects are not reversible in the sense in which the Peltier effect and the thermoelectric effect are reversible, but J. J. Thomson has shown that they are consistent with each other.

302. Measurement of Current.—Instruments which are used to detect the presence of a current, or to measure its strength, by means of the deflection of a magnetic needle, are commonly called *galvanometers.*

The simplest form of the galvanometer is the instrument called the *Schweigger's multiplier.* It consists of a flat spool upon which an insulated wire is wound a number of times. The plane of the coils is vertical, and usually also coincides with the plane of the magnetic meridian. A magnetic needle is suspended in the interior of the spool. When a current is passed through the wire, the needle is deflected from the magnetic meridian. Usually, in order to make the indications of the apparatus more sensitive, a combination of two needles is used. They are joined rigidly together, so that when suspended the lower one hangs in the interior of the spool, and the other in the same plane directly above the spool. These needles are magnetized so that the positive end of one is above the negative end of the other. If they are of nearly equal strength, such a combination will have very little directive tendency in the earth's magnetic field. It is therefore called an *astatic* system. When a current passes in the wire, however, the lines of force due to the current form closed curves passing through the coil, and both needles tend to turn in the same direction. Since the earth's field offers almost no resistance to this tendency, an astatic system will indicate the presence of very feeble currents. The apparatus here described is no longer used to measure currents, but only to detect their presence and direction.

The *tangent galvanometer* is that form of galvanometer which is commonly used to measure electrical currents in electromagnetic units. We will consider it only in one of its simplest forms. In

this form it consists of a circular conductor set up in the earth's magnetic field, so that its plane is parallel with the lines of force, and having a small magnet placed at its centre. The magnet is free to swing in the horizontal plane. If the current i be sent through the circuit, the couple which it will exert on the magnet, on the supposition that the magnet is so short that the force at its poles is the same as that at its centre, is $\frac{2\pi i M}{r}\cos\phi$ (§ 296), where M represents the magnetic moment of the magnet, and ϕ the angle made by the axis of the magnet with the direction of the lines of force. The couple exerted by the field upon the magnet and tending to turn it in the opposite sense is $HM \sin\phi$, where H represents the horizontal intensity of the earth's magnetism. Equilibrium will obtain, and the magnet will be at rest, when these couples are equal, or when $\frac{2\pi i M}{r}\cos\phi = HM\sin\phi$. From this equation we obtain

$$i = \frac{Hr}{2\pi}\tan\phi. \tag{104}$$

The current is therefore proportional to the tangent of the angle of deflection. All the quantities in this expression for current, except H, are either numbers or lengths and may be directly measured; and H may be determined in absolute units (§ 244). The tangent galvanometer therefore permits of the determination of current strength in absolute units.

In the more complicated forms of the instrument, the dimensions and position of its parts are so adjusted that the corrections rendered necessary by the impossibility of fulfilling the conditions assumed in the simple case may be either calculated or avoided.

Weber's *electro-dynamometer* is an instrument with fixed coils like those of the tangent galvanometer, but with a small suspended coil substituted for the magnet. The small coil is usually suspended by the two fine wires through which the current is introduced into it, and the moment of torsion of this so-called *bifilar suspension* enters into the expression for the current strength. The same current is sent through the fixed and the movable coils, and a measure-

ment of its strength can be obtained in absolute units, as with the tangent galvanometer. By a proper series of experiments, this measurement is made independent of the horizontal intensity of the earth's magnetism. When the current is reversed in the instrument, the couple tending to turn the suspended coil does not change. If the effects of terrestrial magnetism can be avoided, the electro-dynamometer can therefore be used to measure rapidly alternating currents.

303. Electromotive Force.—The electromotive force in a circuit is defined as before (§ 273), as the power of establishing or sustaining the conditions which make the expenditure of energy in the circuit possible, and it is measured, as before, by the energy expended in the circuit. If W represent the energy expended in unit time, we have $W = ie$, where e is the electromotive force measured in electromagnetic units. Since we may measure current in electromagnetic units by the tangent galvanometer, this relation enables us to measure electromotive force in absolute electromagnetic units.

The electromagnetic unit of electromotive force is that electromotive force which will produce the expenditure of one unit of energy in unit time, when the electromagnetic unit of current is traversing the conductor.

The dimensions of electromotive force in the electromagnetic system are obtained from the fact that the electromotive force multiplied by the current strength equals the rate of work. The dimensions of rate of work are $\left[\frac{W}{T}\right]$, so that $[e] = \left[\frac{W}{iT}\right] = \frac{ML^2T^{-2}}{M^{\frac{1}{2}}L^{\frac{1}{2}}} = M^{\frac{1}{2}}L^{\frac{3}{2}}T^{-2}$.

In practice another unit is used, called the *volt*. It contains 10^8 C. G. S. electromagnetic units.

It will be shown subsequently (§ 310) that the unit of electromotive force may be defined by the effects produced in a circuit by its motion in a magnetic field, the two definitions being of course consistent with each other.

304. Resistance.—As in the discussion of § 274, we may define the ratio of the electromotive force to the current in any circuit as the resistance in that circuit. The *electromagnetic unit of resistance* is the resistance of that circuit in which unit electromotive force gives rise to unit current, when both these quantities are measured in electromagnetic units.

In practice another unit of resistance is used, called the *ohm.* The *true ohm* contains 10^9 C. G. S. electromagnetic units. The dimensions of resistance in the electromagnetic system are $[r] = \left[\frac{e}{i}\right] = LT^{-1}$.

The standard of resistance, usually called the B. A. unit, determined by the committee of the British Association, has a resistance somewhat less than the true ohm as it is here defined. In practical work resistances are used which have been compared with this standard. The Electrical Congress of 1884 defined the *legal ohm* to be "the resistance of a column of mercury of one square millimetre section and of 106 centimetres of length at the temperature of freezing." This definition has since been modified by increasing the length of the mercury column to 106.3 cm. The legal ohm contains 1.0112 B.A. units. Boxes containing coils of wire of definite resistance, so arranged that by different combinations of them any desired resistance may be introduced into a circuit, are called *resistance boxes* or *rheostats.*

305. Kirchhoff's Laws.—In circuits which are made up of several parts, forming what may be called a network of conductors, there exist relations among the electromotive forces, currents, and resistances in the different branches, which have been stated by Kirchhoff in a way which admits of easy application.

Several conventions are made with regard to the positive and negative directions of currents. In considering the currents meeting at any point, those currents are taken as positive which come up to the point, and those as negative which move away from it. In travelling around any closed portion of the network, those currents are taken as positive which are in the direction of motion, and those

as negative which are opposite to the direction of motion. Further, those electromotive forces are positive which tend to set up positive currents in their respective branches. With these conventions *Kirchhoff's laws* may be stated as follows:

1. The algebraic sum of all the currents meeting at any point of junction of two or more branches is equal to zero. This first law is evident, because, after the current has become steady, there is no accumulation of electricity at the junctions.

2. The sum, taken around any number of branches forming a closed circuit, of the products of the currents in those branches and their respective resistances is equal to the sum of the electromotive forces in those branches. This law can easily be seen to be only a modified statement of Ohm's law.

These laws may be illustrated by their application in a form of apparatus known as *Wheatstone's bridge*. The circuit of the Wheatstone's bridge is made up of six branches. An end of any branch meets two, and only two, ends of other branches, as shown in Fig. 87. In the branch 6 is a voltaic cell with an electromotive force E. In the branch 5 is a galvanometer which will indicate the presence of a current in that branch. In the other branches are conductors, the resistances of which may be called respectively r_1, r_2, r_3, r_4.

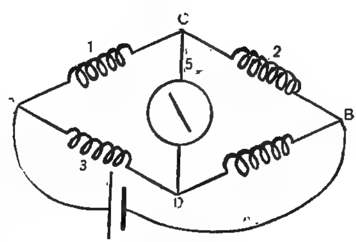

FIG. 87.

From Kirchhoff's first law the sum of the currents meeting at the point C is $i_1 + i_2 + i_5 = 0$, and of those meeting at the point D is $i_3 + i_4 + i_5 = 0$. By the second law, the sum of the products ir in the circuit ADC is $i_1r_1 + i_3r_3 + i_5r_5 = 0$, and in the circuit DBC is $i_4r_4 + i_2r_2 + i_5r_5 = 0$, since there are no electromotive forces in those circuits. If we so arrange the resistances of the branches 1, 2, 3, 4 that the galvanometer shows no deflection, the current i_5 is zero, and these equations give the relations $i_1 = -i_2$, $i_3 = -i_4$, $i_1r_1 = -i_3r_3$, $i_2r_2 = -i_4r_4$. From these four equations

follows at once a relation between the resistances, expressed in the equation

$$r_1 r_4 = r_2 r_3. \tag{105}$$

If, therefore, we know the value of r_3 and know the ratio of r_1 to r_2, we may obtain the value of r_4.

This method of comparing resistances by means of the Wheatstone's bridge is of great importance in practice. By the use of a form of apparatus known as the *British Association bridge* the method can be carried to a high degree of accuracy. In this form of the bridge, the portion marked ACB (Fig. 87) is a straight cylindrical wire, along which the end of the branch CD is moved until a point C is found, such that the galvanometer shows no deflection. The two portions of the wire between C and A, and C and B, are then the two conductors of which the resistances are r_1 and r_2, and these resistances are proportional to the lengths of those portions (§ 275). The ratio of r_1 to r_2 is therefore the ratio of the lengths of wire on either side of C, and only the resistance of r_3 need be known in order to obtain that of r_4.

It is often convenient in determining the relations of current and resistance in a network of conductors to use Ohm's law directly, and consider the difference of potential between the two points on a conductor as equal to the product ir. When a part of a circuit is made up of several portions which all meet at two points A and B, the relation between the whole resistance and that of the separate parts may be obtained easily in this way. For convenience in illustration we will suppose the divided circuit (Fig. 88) made up of only three portions, 1, 2, 3, meeting at the points A and B, and that no electromotive force exists in those portions. Then the difference of potential between A and B is $V_A - V_B = i_1 r_1 = i_2 r_2 = i_3 r_3$. We have also by Kirchhoff's first law $-i_4 = i_1 + i_2 + i_3$. By the combination of these equations we obtain

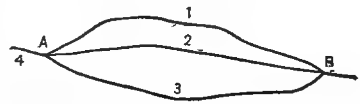

FIG. 88.

$$-i_4 = (V_A - V_B)\left(\frac{1}{r_1} + \frac{1}{r_2} + \frac{1}{r_3}\right). \tag{106}$$

The current in the divided circuit equals the difference of potential between A and B multiplied by the sum of the reciprocals of the resistances of the separate portions. If we set this sum equal to $\frac{1}{r}$, and call r the resistance of the divided circuit, we may say that the reciprocal of the resistance of a divided circuit is equal to the sum of the reciprocals of the resistances of the separate portions of the circuit. When there are only two portions into which the circuit is divided, one of them is usually called a *shunt,* and the circuit a shunt circuit.

The rules for joining up sets of voltaic cells in circuits so as to accomplish any desired purpose may be discussed by the same method. Let us suppose that there are n cells, each with an electromotive force e and an internal resistance r, and that the external resistance of the circuit is s. If m be a factor of n, and if we join up the cells with the external resistance so as to form a divided circuit of m parallel branches, each containing $\frac{n}{m}$ cells, we shall have for the electromotive force in such a circuit $\frac{ne}{m}$, and for the resistance of the circuit $s + \frac{nr}{m^2}$. The current in the circuit is therefore $i = \frac{mne}{m^2s + nr}$. Two cases may arise which are common in practice. The resistance s of the external circuit may be so great that, in comparison with m^2s, nr may be neglected. In that case i is a maximum when $m = 1$, that is, when the cells are arranged *tandem,* or in *series,* with their unlike poles connected. On the other hand, if m^2s be very small as compared with nr, it may be neglected, and i becomes a maximum when $m = n$, that is, when the cells are arranged *abreast,* or in *multiple arc,* with their like poles in contact.

306. Induced Currents.—It was shown in § 277 that the movement of a magnet in the neighborhood of a closed circuit will give rise, in general, to an electromotive force in the circuit, and that

the current due to this electromotive force will be in the direction opposite to that current which, by its action upon the magnet, would assist the actual motion of the magnet. This current is called an *induced current.* From the equivalence between a magnetic shell and an electrical current, it is plain that a similar induced current will be produced in a closed circuit by the movement near it of an electrical current or any part of one. Since the joining up or breaking the circuit carrying a current is equivalent to bringing up that same current from an infinite distance, or removing it to an infinite distance, it is further evident that similar induced currents will be produced in a closed circuit when a circuit is made or broken in its presence.

The demonstration of the production of induced currents in § 277 depends upon the assumption that the path of the magnet pole is such that work is done upon it by the current assumed to exist in the circuit. The potential of the magnet pole relative to the current is changed.

The change in potential from one point to another in the magnetic field due to a closed current is (§ 290) $i(\Omega' - \Omega + 4\pi n)$, and the work done on a magnet pole m, in moving it from one point to another, is $mi(\Omega' - \Omega + 4\pi n)$. In the demonstration of § 277 we may substitute $m(\Omega' - \Omega + 4\pi n)$ for A, and, provided the change in potential be uniform, we obtain at once the expression $-\dfrac{m(\Omega' - \Omega + 4\pi n)}{t}$ for the electromotive force due to the movement of the magnet pole. If the change in potential be not uniform, we may conceive the time in which it occurs to be divided into indefinitely small intervals, during any one of which, t, it may be considered uniform. Then the limit of the expression $-\dfrac{m(\Omega' - \Omega + 4\pi n)}{t}$, as t becomes indefinitely small, is the electromotive force during that interval.

The current strength due to this electromotive force is

$$i' = -\frac{m(\Omega' - \Omega + 4\pi n)}{rt}.$$

If the induced current be steady, the total quantity of electricity flowing in the circuit is expressed by $i't = -\frac{m(\Omega' - \Omega + 4\pi n)}{r}$.

The total quantity of electricity flowing in the circuit depends, therefore, only upon the initial and final positions of the magnet pole, and the number of times it passes through the circuit, and not upon its rate of motion. The electromotive force due to the movement of the magnet, and consequently the current strength, depends, on the other hand, upon the rate at which the potential changes with respect to time.

A more general statement of the mode in which induced currents are produced may be given in terms of the changes in the number of tubes of induction which pass through the circuit. When the number of tubes of induction which pass through a circuit is altered, an electromotive force is induced in the circuit which is proportional to the rate of change of the number of tubes of induction. This law may be easily proved, as in the special case already considered, if the change in the number of tubes of induction be produced by a movement of magnet poles or their equivalents, and not by changes in other currents in the field; in case there are other currents in the field, the interactions between them introduces conditions which cannot be discussed by elementary methods. The law, however, is a perfectly general one, and holds for all cases in which the tubes of induction passing through the circuit change in number.

While we cannot, by elementary methods, determine exactly the laws of the production of an induced current in a circuit by changes in the currents in neighboring circuits, we may yet form some idea of the induced current by considering the magnetic field about the circuits. Suppose that a current traverses circuit 1 and that there is no current in circuit 2; circuit 2 encloses a number of tubes of induction due to the current in circuit 1. If the current in circuit 1 be suddenly interrupted, these tubes of induction are removed from circuit 2, and from the dynamical principle that a change is resisted by the non-conservative forces to which it gives rise, there will arise in circuit 2 a current tending to maintain the tubes

within it. If the two circuits are parallel, this current will be in the same sense as that in circuit 1. The current induced in circuit 2 gives rise to tubes of induction which enter circuit 1, and their entrance into circuit 1 is resisted by a current tending to repel them from circuit 1, or to set up tubes of induction in the opposite sense. Thus there will be a small current in circuit 1 in the opposite sense to that originally in it and the current in circuit 1 will therefore diminish more rapidly than if circuit 2 were not present. On the other hand, if neither circuit carries a current, and a current be suddenly impressed on circuit 1, the tubes of induction to which it gives rise will enter circuit 2, and will be resisted by a momentary current in circuit 2 tending to repel them, or to set up tubes of induction in the opposite sense. Thus the induced current in circuit 2 in this case, if the two circuits are parallel, is in the opposite sense to that in circuit 1. This current in circuit 2 will in turn set up tubes of induction which enter circuit 1 and are there resisted by a momentary small current which will be in the same sense as that impressed upon circuit 1. Thus the presence of circuit 2 will temporarily increase the current in circuit 1.

The fact that induced currents are produced in a closed circuit by a variation in the number of lines of magnetic force included in it was first shown experimentally by Faraday in 1831. He placed one wire coil, in circuit with a voltaic battery, inside another which was joined with a sensitive galvanometer. The first he called the *primary*, the second the *secondary*, circuit. When the battery circuit was made or broken, deflections of the galvanometer were observed. These were in such a direction as to indicate a current in the secondary coil, when the primary circuit was made, in the opposite direction to that in the primary, and when the primary circuit was broken, in the same direction as that in the primary. When the positive pole of a bar magnet was thrust into or withdrawn from the secondary coil, the galvanometer was deflected. The currents indicated were related to the direction of motion of the positive magnet pole, as the directions of rotation and propulsion in a left-handed screw. The direction of the induced currents

in these experiments is easily seen to be in accordance with the law above stated. A simple statement, known as *Lenz's law*, which enables us to determine the sense of an induced current produced by the motion of a magnet or a circuit, is as follows: When an induced current is produced, it is always in such a sense as to oppose the action which produces it. This is equivalent to the statement that the induced current tends to oppose the change in the number of tubes of induction which pass through the circuit.

The case in which an induced current in the secondary circuit is set up by making the primary circuit is, as has been said, an extreme case of the movement of the primary circuit from an infinite distance into the presence of the secondary. The experiments of Faraday and others show that the total quantity of electricity induced when the primary circuit is made is exactly equal and opposite to that induced when the primary circuit is broken. They also show that the electromotive force induced in the secondary circuit is independent of the materials constituting either circuit, and is proportional to the current strength in the primary circuit. These results are consistent with the formula already deduced for the induced current.

307. Currents of Self-induction.—If the current in a circuit be changed, the number of tubes of induction which pass through the circuit will vary, and an induced current will be set up in the circuit. If there be originally no current in the circuit and if an electromotive force be suddenly impressed upon it, so that the current which finally exists in the circuit is i, the number of tubes of induction developed through the circuit is equal to Li (§ 293). Let t be the time required for the current to rise to its full value; then the average electromotive force induced in the circuit by the increase in the number of tubes of induction which pass through it will be $\frac{Li}{t}$, and the average current will be $\frac{Li}{rt}$. The total current due to this induced electromotive force is therefore $\frac{Li}{r}$, and is opposed, in sense, to the current impressed upon the circuit. If the circuit be suddenly broken, the same expression represents the total

induced current due to the loss of the tubes of induction which pass through the circuit; this current is in the same sense as the current of the circuit. Since by Ohm's law $i = \frac{e}{r}$, where e is the electromotive force impressed upon the circuit, the average electromotive force is in both these cases $\frac{eL}{rt}$. Now t, the time required for the current to rise from zero to its full value, or to sink from its full value to zero, is very small, and the average electromotive force of induction may be much larger than the electromotive force of the circuit. When the current is made, this induced electromotive force diminishes the electromotive force of the circuit; so that the current is established gradually and not instantaneously. The time required to establish the current depends upon the resistance and self-induction of the circuit. When the circuit is broken, the electromotive force of induction is in the same sense as that of the circuit, and produces a momentary current which is much greater than the steady current of the circuit. The induced electromotive force is frequently so high as to cause the current to leap across the gap formed where the circuit is broken, and to give rise to a spark at that gap. The induced current thus formed is often called the *extra current* or the *current of self-induction.* It should be noted that the induced electromotive force is proportional to the coefficient of self-induction of the circuit. The establishment of a current in the circuit may therefore be retarded and the extra current at the break may be increased by so arranging the circuit as to increase its coefficient of self-induction; while by so winding the circuit that its coefficient of self-induction is reduced to a minimum these effects may be almost entirely avoided. A wire doubled on itself, and coiled so that a current in it always passes in opposite directions through immediately contiguous portions of the wire, will manifestly have a very small coefficient of self-induction; such a coil is called a *non-inductive coil.*

308. Alternating Currents.—If the electromotive force in a circuit be made to vary, especially if it be made to change in sense, the tubes of induction which pass through the circuit will also vary,

and the current in the circuit will vary in a way dependent not only on the variations in the electromotive force, but also on the currents produced by induction. The case of the greatest interest and importance is that in which the electromotive force varies periodically; in this case the current also varies periodically. It may be shown, by a method which cannot be given here, that the maximum value of the current is never as great as that deduced from the maximum electromotive force on the supposition that the current follows Ohm's law. The formula which expresses the maximum value of the current is $\frac{e}{\left(r^2 + \frac{4\pi^2 L^2}{T^2}\right)^{\frac{1}{2}}}$, where e is the maximum electromotive force and T the period of the alternation. The denominator of this expression is a quantity of the same order as resistance, but it involves, besides the resistance of the circuit, its coefficient of self-induction and the period. In case $\frac{4\pi^2 L^2}{T^2}$ is very large in comparison with r^2, the current has its maximum value at the time when the electromotive force is zero, and is zero when the electromotive force is a maximum. The theory further shows that the rate of propagation of the electrical disturbance along the conductor is a function of the period of the alternation, being less when the period is greater. When the period is infinitesimal, or in general when it is very small, the velocity is equal to the velocity v, the ratio between the electrostatic and the electromagnetic units (§ 311), or to the velocity of light. The currents developed in the conductor, by rapid alternations of electromotive force, are not the same for all parts of the cross-section of the conductor, but diminish from the outside of the conductor inwards. For very rapid alternations the currents exist only in a small layer near the surface of the conductor. These deductions of theory have been fully confirmed by experiment.

309. Apparatus employing Induced Currents.—The production of induced currents by the relative movements of conductors and magnets is taken advantage of in the construction of pieces of

apparatus which are of great importance not only for laboratory use but in the arts.

The *telephonic receiver* consists essentially of a bar magnet around one end of which is carried a coil of fine insulated wire. In front of this coil is placed a thin plate of soft iron. When the coils of two such instruments are joined in circuit by conducting wires, any disturbance of the iron diaphragm in front of one coil will change the magnetic field near it, and a current will be set up in the circuit. The strength of the magnet in the other instrument will be altered by this current, and the diaphragm in front of it will move. When the diaphragm of the first instrument, or *transmitter*, is set in motion by sound waves due to the voice, the induced currents, and the consequent movements of the diaphragm of the second instrument, or *receiver*, are such that the words spoken into the one can be recognized by a listener at the other.

Other transmitters are generally used, in which the diaphragm presses upon a small button of carbon. A current is passed from a battery through the diaphragm, the carbon button, and the rest of the circuit, including the receiver. When the diaphragm moves, it presses upon the carbon button, and alters the resistance of the circuit at the point of contact. This change in resistance gives rise to a change in the current, and the diaphragm of the receiver is moved. The telephone serves in the laboratory as a most delicate means of detecting rapid changes of current in a circuit.

The various forms of magneto-electrical and dynamo-electrical machines are too numerous and too complicated for description. In all of them an arrangement of conductors, usually called the *armature*, is moved in a powerful magnetic field, and a suitable arrangement is made by which the currents thus induced may be led off and utilized in an outside circuit. The magnetic field is sometimes established by permanent magnets, and the machine is called a *magneto-machine*. In most cases, however, the circuit containing the armature also contains the coils of the electromagnets to which the magnetic field is due. When the armature rotates, a current starts in it, at first due to the residual magnetism of some part of

the machine: this current passes through the field magnets and increases the strength of the magnetic field. This in turn reacts upon the armature, and the current rapidly increases until it attains a maximum due to the fact that the magnetic field does not increase proportionally to the current which produces it. Such a machine is called a *dynamo-machine.* By suitable arrangements of the conductors which lead the current from the machine, either direct or alternating currents may be obtained.

The *induction coil*, or *Ruhmkorff's coil*, consists of two circuits wound on two concentric cylindrical spools. The inner or primary circuit is made up of a comparatively few layers of large wire, and the outer, or secondary, of a great number of turns of fine wire. Within the primary circuit is a bundle of iron wires, which, by its magnetic action, increases the electromotive force of the induced current in the secondary coil. Some device is employed by which the primary circuit can be made or broken mechanically. The electromotive force of the induced current is proportional to the number of windings in the secondary coil, and as this is very great the electromotive force of the induced current greatly exceeds that of the primary current. The electromotive force of the induced current set up when the primary circuit is broken is further heightened by a device proposed by Fizeau. To two points in the primary circuit, one on either side of the point where the circuit is broken, are joined the two surfaces of a condenser. When the circuit is broken, the extra current, if the condenser be not introduced, forms a long spark across the gap, and so prolongs the fall of the primary current to zero. The electromotive force of the induced current is therefore not so great as it would be if the fall of the primary current could be made more rapid. When the condenser is introduced, the extra current is partly spent in charging the condenser, the difference of potential between the two sides of the gap is not so great, the length of the spark and consequently the time taken by the primary current to become zero is lessened, and the electromotive force of the induced current is proportionally increased.

310. Determination of the Unit of Resistance.—If the circuit considered in § 306 move from a point where its potential relative to the magnet pole is $m\Omega'$ to one where it is $m\Omega$, provided that the magnetic pole do not pass through the circuit, and that the movement be so carried out that the induced current is constant, the electromotive force of the induced current is $-\frac{m(\Omega' - \Omega)}{t}$. If the movement take place in unit time, and if $m(\Omega' - \Omega)$ also equal unity, the electromotive force in the circuit is the *unit electromotive force.*

The expression $m(\Omega' - \Omega)$ is equivalent to the change in the number of tubes of induction passing through the circuit in the positive direction. More generally, then, if a circuit or part of a circuit so move in a magnetic field that, in unit time, the number of tubes of induction passing through the circuit in the positive direction increase or diminish by unity, at a uniform rate, the electromotive force induced is unit electromotive force.

This definition is consistent with the one given in § 303. For, the energy of a circuit carrying the current i, due to the field in which it is placed, equals iN, and the change of this energy in unit time is the energy expended in the circuit in that time. But this change in energy is $i\,\frac{N' - N}{t}$, and $\frac{N' - N}{t}$ is the electromotive force, so that ie represents the energy expended in unit time.

A simple way in which the problem can be presented is as follows: Suppose two parallel straight conductors at unit distance apart, joined at one end by a fixed cross-piece. Suppose the circuit to be completed by a straight cross-piece of unit length which can slide freely on the two long conductors. Suppose this system placed in a magnetic field of unit intensity, so that the lines of force are everywhere perpendicular to the plane of the conductors. Then, if we suppose the sliding piece to be moved with unit velocity perpendicular to itself along the parallel conductors, the electromotive force set up in the circuit will be the unit electromotive force, and if it move with any other velocity v, the electromotive force will be equal to v.

If we now insert a galvanometer in the fixed cross-piece, and suppose the resistance of all the circuit except the sliding piece to be negligible, and move the sliding piece at such a rate that the current in the galvanometer is unity, we have the resistance of the sliding piece determined from the velocity with which it moves. For, by Ohm's law, $i = \frac{e}{r}$, and since $i = 1$ and $e = v$, we have $r = v$.

Such an arrangement as that here described is of course impossible in practice, but it embodies the principle of the method actually used to determine the unit of resistance by the Committee of the British Association. In their method, a circular coil of wire, in the centre of which was suspended a small magnetic needle, was mounted so as to rotate with constant velocity about a vertical diameter. From the dimensions and velocity of rotation of the coil and the intensity of the earth's magnetic field, the induced electromotive force in the coil was calculated. The current in the same coil was determined by the deflection of the small magnet. The ratio of these two quantities gave the resistance of the coil.

311. Ratio between the Electrostatic and Electromagnetic Units.—When the dimensions of any electrical quantity derived from its electrostatic definition are compared with its dimensions derived from its electromagnetic definition, the ratio between them is always of the dimensions of some power of a velocity. The ratio between the electrostatic and electromagnetic unit of any electrical quantity is, therefore, some power of a velocity. If this ratio be obtained for any set of units, the number expressing it will also express some power of a velocity. This velocity is an absolute quantity or constant of Nature. Whatever changes are made in the units of length and time, the number expressing this velocity in the new units will also express the ratio of the two sets of electrical units.

This ratio, which is called v, can be measured in several ways.

The method which was first used, by Weber and Kohlrausch, depends upon the comparison of a quantity of electricity measured

in the two systems. From the dimensions of current in the electromagnetic system we have the dimensions of quantity $[q] = [iT] = M^{\frac{1}{2}}L^{\frac{1}{2}}$. The dimensions of quantity in the electrostatic system are $[Q] = M^{\frac{1}{2}}L^{\frac{3}{2}}T^{-1}$. The ratio of these dimensions is $\left[\frac{Q}{q}\right] = LT^{-1}$, or, the number of electrostatic units of quantity in one electromagnetic unit is the velocity v.

In Weber and Kohlrausch's method the charge of a Leyden jar was measured in electrostatic units by a determination of its capacity and the difference of potential between its coatings. The current produced by its discharge through a galvanometer was used to measure the same quantity in electromagnetic measure.

Thomson determined v by a comparison of an electromotive force measured in the two systems. He sent a current through a coil of very high known resistance, and measured it by an electrodynamometer. The electromagnetic difference of potential between the two ends of the resistance coil was then equal to the product of the current by the resistance. The electrostatic difference of potential between the same two points was measured by an absolute electrometer. From the dimensional formulas we have $\left[\frac{E}{e}\right] = \frac{M^{\frac{1}{2}}L^{\frac{1}{2}}T^{-1}}{M^{\frac{1}{2}}L^{\frac{3}{2}}T^{-2}} = L^{-1}T$. The number of electromagnetic units of electromotive force in one electrostatic unit is v. The ratio of the numbers expressing the electromagnetic and the electrostatic measures of the electromotive force in Thomson's experiment is therefore the quantity v. This experiment was carried out by Maxwell in a different form, in which the electrostatic repulsion of two similarly charged disks was balanced by an electromagnetic attraction between currents passing through flat coils on the back of the two disks.

Other methods, depending on comparisons of currents, of resistances, and other electrical quantities, have been employed. The methods described are historically interesting as being the first ones used. The values of v obtained by them differed rather widely from one another. Recent determinations, however, give

more consistent results. It is found that v, considered as a velocity, is about 3.10^{10} centimetres in a second. This velocity agrees very closely with the velocity of light.

An experiment was executed by Rowland in which this velocity v was obtained by comparison with the actual velocity of a moving charge. The principle of the experiment is as follows: If we consider an indefinitely extended plane surface on which the surface density of electrification is σ measured in electrostatic units, or $\frac{\sigma}{v}$ measured in electromagnetic units, since the ratio of the electrostatic to the electromagnetic unit of quantity is v; and conceive it to move in its own plane with a velocity x; the charge moving with it may be considered as the equivalent of a current in that surface, the strength of which, measured by the quantity of electricity which crosses a line of unit length, perpendicular to the direction of movement, in unit time, is $\frac{\sigma x}{v}$. The force due to such a current on a magnet may be calculated. Conversely, if the force on the magnet be observed, and the surface density σ and the velocity x be also measured, the value of v may be calculated. The probability of such an action as the one here described was stated by Maxwell.

The experiment by which Rowland verified Maxwell's view consisted in rotating a disk cut into numerous sectors, each of which was electrified, under an astatic magnetic needle. During the rotation of the disk, a deflection of the needle was observed, in the same sense as that in which it would have moved if a current had been flowing about the disk in the direction of its rotation. From the measured values of the deflecting force, of the surface density of electrification on the disk, and the velocity of rotation, Rowland calculated a value of v which lies between those given by Weber and Maxwell.

312. Oscillatory Discharge of a Condenser.—If a condenser be discharged through a circuit, the current in the circuit will manifestly depend on the original difference of potential between the

plates of the condenser, on the resistance of the circuit, and on its self-induction. In case the resistance of the conductor is greater than a certain value, determined by the self-induction of the circuit and the capacity of the condenser, the current will decrease steadily from its value at the beginning of the discharge to zero. But in most cases this condition is not fulfilled, and in these cases the current assumes another character. It goes through a series of alternations in opposite senses, which are periodic in the sense that the successive maximum values of the current follow each other at equal intervals of time, though the absolute values of these maxima diminish very rapidly. The discharge in this case is called an *oscillatory discharge.* That the discharge of an ordinary condenser is of this nature was discovered by Joseph Henry, from the manner in which needles were magnetized by the discharge passed through small coils of wire. The theory was afterwards indicated by William Thomson, and his conclusions were fully confirmed by the investigations of Feddersen. Feddersen observed the spark produced by the discharge in a rotating mirror, and found that instead of giving a single line of light in the mirror, it gave a series of lines at equal distances apart. He showed that the period of the oscillation could be changed by changing the conditions of the circuit, and that by sufficiently increasing the resistance without correspondingly increasing the self-induction, the period of the oscillations was increased till finally the discharge ceased to exhibit any oscillations whatever.

313. Electromagnetic Waves.—According to Maxwell's theory of electricity, an oscillatory discharge of the sort just described ought to set up a series of disturbances in the medium surrounding the circuit, which proceed outward from the circuit in the manner of waves set up in any medium by a disturbance at a point in it; such disturbances may be called *electromagnetic waves.* The existence of such waves was demonstrated by Hertz, and the examination of their properties by him and by others has shown that they conform practically in all respects to the predictions of the theory. The arrangement used by Hertz to set up electromagnetic waves, called

by him the *vibrator*, consisted, in a typical form, of two metal plates set up in the same plane, each carrying a rod on the end of which was a small sphere; the plates were so placed that the spheres were near together, with a short air-gap left between them. When the plates were joined to the two terminals of a Holtz machine or to the two terminals of an induction-coil, a series of sparks passed across the air-gap between the spheres. It may be shown that the electrical oscillations in the sparks are practically independent of the peculiarities of the Holtz machine or the induction-coil, and depend only on the capacity of the plates and the resistance and self-induction of the plates and rods carrying the spheres. The electromagnetic waves originate at the spark-gap. The instrument used by Hertz to detect the waves, called by him the *resonator*, consisted simply of a plane circuit broken at one point by a very small gap; the presence of an electrical disturbance in this circuit could be detected by the appearance of sparks in the gap. The dimensions of the resonator were so adjusted that the period of the electrical oscillations which would originate in it if a momentary discharge were sent through it was the same as the period of the discharge of the vibrator. The electromagnetic waves coming from the vibrator set up electrical disturbances in the resonator, which were detected by the passage of sparks across the gap.

By the aid of these instruments, Hertz first proved the existence of electromagnetic waves; he then impressed upon a wire an electrical oscillation of the same period as that sent through the air around the wire, and compared the rate of propagation of the two disturbances. Hertz's own experiments were misleading, for reasons which perhaps cannot now be given, but Sarasin and de la Rive, working under more favorable circumstances, reached the conclusion, which was accepted by Hertz, that the velocity of propagation of the wave in the wire was the same as that in air, when the periods of vibrations were very small. This result is in accordance with theory. It follows immediately from the view we have taken that the current is due to the movement of tubes of force through the dielectric surrounding the circuit.

Hertz showed that the electromagnetic waves were reflected from metal surfaces according to the law for the reflection of light (§ 333), and that nodal points or points of interference between the waves advancing from the vibrator and those returning from the mirror could be detected, and thus the wave-length of the disturbance determined. If the wave-length and the period be known, the velocity of the wave may be calculated; an approximate calculation of the period was made from the dimensions of the vibrator, and the velocity of the waves determined to be of the same order of magnitude as the velocity of light. Subsequent experiments, under more favorable conditions and with vibrators which permit a more precise calculation of the period, have confirmed the conclusion of theory, that the velocity of very short electromagnetic waves is the same as the velocity of light.

Hertz also proved that the electromagnetic waves are refracted (§ 334) when they pass from one medium into another. By the use of a large prism of pitch he obtained a considerable deviation of the waves and was able to calculate the index of refraction of pitch for such waves; he obtained a number of the same order of magnitude as the index of refraction of ordinary refracting bodies for light.

Owing to the way in which these waves are generated by an oscillatory discharge in one line, the waves which proceed from them are polarized (§ 376), that is, the electromotive forces transmitted through the air have always the same direction. Hertz interposed in the path of the waves a screen made of a number of parallel wires; he found that when the wires were parallel with the line of the discharge or with the electromotive forces in the successive waves, the waves were almost entirely absorbed by the wires. If, on the other hand, the wires were set so as to be at right angles to the electromotive forces in the waves, the waves passed through the screen without modification. The screen therefore exhibits a property analogous to that of tourmaline in polarized light (§ 379). Righi and others have observed similar effects produced by the interposition of blocks of wood in the path of the waves, which

absorbed the waves in different degrees according as the grain of the wood was parallel with the electromotive forces of the waves or was transverse to them.

Trouton observed that, when the electromagnetic waves were directed obliquely against a thick stone wall, the waves were, in general, partly reflected and partly transmitted. The ratio between the intensity of the reflected and transmitted waves depended upon the obliquity and upon the angle between the direction of the electromotive forces in the waves and a plane containing the normal to the incident waves and the normal to the reflecting surface. For a certain obliquity, the incident waves were entirely reflected, in case the electromotive forces in the waves were at right angles to the plane of incidence, or were parallel with the reflecting surface. In this case there was no transmitted wave. When, with the same obliquity, the electromotive forces in the waves were in the plane of incidence, there was no reflected wave and the incident wave was entirely transmitted. These properties are exactly analogous to those exhibited by the reflection and refraction of polarized light (§ 377). Trouton found that he could not obtain similar action from sheets of window-glass. These laws of reflection and refraction, and the impossibility of obtaining reflections and refractions consistent with them when the wave-length is long in comparison with the thickness of the reflecting body, are consistent with theory.

Several observers have determined the velocity of the electromagnetic waves in various dielectrics in comparison with their velocity in air. According to Maxwell's theory the ratio of the velocity in air to the velocity in the dielectric, or the index of refraction of the dielectric (§ 334), is equal to the square root of the dielectric constant. This conclusion of theory has been verified in very many cases.

The consideration of these experiments will be resumed in connection with the electromagnetic theory of light.

CHAPTER VI.

THERMO-ELECTRIC RELATIONS OF THE CURRENT.

314. Thermo-electric Currents.—The heating or cooling of a junction of two dissimilar metals by the passage of a current, referred to in § 270 as the Peltier effect, is the reverse of a phenomenon discovered in 1822–23 by Seebeck. He found that, when the junction of two dissimilar metals was heated, a current was sent through any circuit of which they formed a part. It has since been shown that the same phenomenon appears if the junction of two liquids, or of a liquid and a metal, be heated. This fact, as has been already shown in § 277, follows as a result of the Peltier phenomenon. If we designate by P the heat developed at the junction by the passage of unit current for unit time, we may substitute it for the expression $\frac{A}{t}$ in the general equation of § 277, and obtain $I = \frac{E - P}{R}$. The counter-electromotive force set up at the heated junction is the coefficient P.

If the electromotive force E and the current I be reversed in the circuit, the junction is cooled and we obtain $I = \frac{E + P}{R}$. The electromotive force at the junction, therefore, tends to increase the electromotive force of the circuit. Since the current in this case is opposite to the current in the case in which the junction is heated, the direction of the electromotive force at the junction is the same in both cases. If there be no electromotive force E in the circuit, we have $I = -\frac{P}{R}$ in case a unit of heat is communicated to the

junction and absorbed by it in unit time, and $I = \frac{P}{R}$ in case a similar quantity of heat is removed from the junction by cooling.

If two strips of dissimilar metals, for example antimony and bismuth, be placed side by side, and united at one end of the pair, being everywhere else insulated from one another, the combination is called a *thermoelectric element*. If several such elements be joined in series, so that their alternate junctions lie near together and in one plane, as indicated in Fig. 89, such an arrangement is called a *thermopile*. When one face of the pile is heated, the electromotive force of the pile is the sum of the electromotive forces of the several elements. Such an instrument was used by Melloni, in connection with a delicate galvanometer, in his researches on radiant heat.

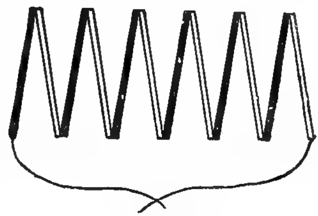

FIG. 89.

When a thermoelectric element is constructed of any two metals, that metal is said to be *thermoelectrically positive* to the other from which the current flows across the heated junction.

315. Thermoelectric Series.—It was found by the experiments of Seebeck himself, and those of others, that the metals may be arranged in a series such that any metal in it is thermoelectrically positive to those which follow it, and thermoelectrically negative to those which precede it.

If a circuit be formed of any two metals in this series, and one of the junctions be kept at the temperature zero, while the other is heated to a fixed temperature, there will be set up an electromotive force which can be measured. If now the circuit be broken at either junction, and the gap filled by the introduction of any other metals of the series, then, provided that the junction which has not been disturbed be kept at the temperature which it previously had, and that the other junctions in the circuit be all raised to the temperature of the junction which was broken, there will be the same electromotive force in the circuit as existed before the introduction

of the other metals of the series. It is manifest, then, that in a circuit made up of any metals whatever, at one temperature, no electromotive force can be set up by changing the temperature of the circuit as a whole.

Thomson showed that it is not necessary for the production of thermal currents that the circuit should contain two metals; but that want of homogeneity arising from any strain of one part of an otherwise homogeneous circuit will also admit of the production of such currents. It has also been shown that when a portion of an iron wire is magnetized, and is heated near one of the poles produced, a thermal current will be set up.

Cumming discovered in 1823 that, if the temperature of one junction of a circuit of two metals be gradually raised, the current produced will increase to a maximum, then decrease until it becomes zero, after which it is reversed and flows in the opposite direction. The experiments of Avenarius, Tait, and Le Roux show that, for almost all metals, the temperature of the hot junction at which the maximum current occurs is the mean between the temperatures of the two junctions at which the current is reversed.

316. Thermoelectric Diagram.—The facts hitherto discovered in relation to thermoelectricity may be collected in a general formula or exhibited by means of a thermoelectric diagram.

Let us consider a circuit of two metals, copper and lead, in which both junctions are at first at the same temperature. We may assume that there is an equal electromotive force at both junctions acting from lead to copper. If one of the junctions be gradually heated, a current will be set up, passing from lead to copper across the hot junction. The heating has disturbed the equilibrium of electromotive forces, and has increased the electromotive force across the hot junction from lead to copper. The rate at which this electromotive force changes with change in the temperature is called the *thermoelectric power* of the two metals. That is, if E represent the electromotive force, t the temperature, and θ the thermoelectric power, we have $\frac{E_1 - E_0}{t_1 - t_0} = \theta_1$, in the limit

where t_1 and t_0 are indefinitely near one another. Hence if we lay off on the axis of abscissas (Fig. 90) an infinitesimal length $t_1 - t_0$, and erect as ordinate the corresponding thermoelectric power θ_1, the area of the rectangle formed by the two lines will represent the electromotive force $E_1 - E_0$, due to the change in temperature. If, beginning at the point t_1, we lay off the similar infinitesimal length $t_2 - t_1$, and erect as ordinate the thermoelectric power θ_2, we shall obtain another rectangle representing the electromotive force $E_2 - E_1$. So for any temperature changes the total area of the figure bounded by the axis of temperatures, by the ordinates representing the thermoelectric powers at the temperatures t_0 and t_x, and by the curve AA' passing through the summits of the rectangles so obtained, will represent the electromotive force due to the heating of the junction from t_0 to t_x.

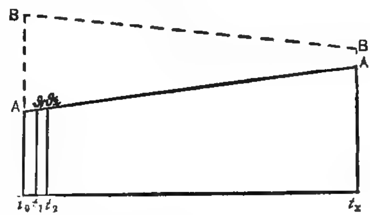

FIG. 90.

It was found by Tait and Le Roux that the thermoelectric power, referred to lead as a standard, of all metals but iron and nickel, is proportional to the rise in temperature. The curve AA' is therefore for those metals a straight line. For iron and nickel the curve is not straight.

For another metal in comparison with lead, the line BB', corresponding to the line AA' for copper, may have a different direction. From what has been said about the possibility of arranging the metals in a thermoelectric series, it is evident that the thermoelectric power between copper and the other metal is the difference of their thermoelectric powers referred to lead, and that the electromotive force at the junction of the two metals, due to a rise of temperature from t_0 to t_x, is represented by the area of the figure contained by the two terminal ordinates and the two lines AA' and BB'. The thermoelectric power is reckoned positive when the current sets from lead to copper across the hot junction. In the diagram the thermoelectric power AB is positive, and the electro-

motive force indicated by the area is from copper to the other metal across the hot junction. At the point where the lines AA' and BB' intersect, the thermoelectric power for the two metals vanishes. The temperature at which this occurs is called the *neutral temperature,* and is designated by t_n. When the temperature t_x lies on the other side of the neutral temperature from t_0, the thermoelectric power becomes negative, and the electromotive force due to the rise in temperature from t_n to t_x is negative. In Fig. 91 it is at once seen that $A'B'$ is negative for t_x, and that the area $NA'B'$ is also negative. The electromotive force due to a rise of temperature from t_0 increases until the temperature of the hot junction is t_n, when it is a maximum, and then decreases. When the area $NA'B'$ becomes equal to the area ANB, the total electromotive force is zero; when $NA'B'$ is greater than ANB, the electromotive force becomes negative, and the current is reversed. In case AA' and BB' are straight lines, it is plain that the temperature t_x, at which this reversal occurs, will be such that the neutral temperature t_n is a mean between t_0 and t_x.

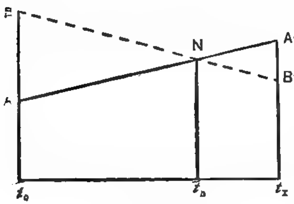

FIG. 91.

The same facts can be represented by a general formula. Thomson first pointed out that the fact of thermoelectric inversion necessitates the view that the thermoelectric power at a junction is a function of the temperature of that junction. Avenarius embodied this idea in a formula, which his own researches, and those of Tait, show to be closely in agreement with experiment. Let us call the hot junction 1 and the cool junction 2, and set the electromotive force at each junction as a quadratic function of the absolute temperatures. We have $E_1 = A + bt_1 + ct_1^2$ and $E_2 = A + bt_2 + ct_2^2$, where A, b, and c are constants. The difference $E_1 - E_2$, or the electromotive force in the circuit, is $E_1 - E_2 = b(t_1 - t_2) + c(t_1^2 - t_2^2) = (t_1 - t_2)\big(b + c(t_1 + t_2)\big)$.

This equation may be put in the form used by Tait, if we write $b = at_n$ and $c = -\frac{a}{2}$. We then have

$$E_1 - E_2 = a(t_1 - t_2)\big(t_n - \tfrac{1}{2}(t_1 + t_2)\big). \tag{107}$$

The electromotive force in the circuit can become zero when either of these terms equals zero. It has been already stated that when $t_1 = t_2$, or when both junctions are at the same temperature, there is no electromotive force in the circuit. When $\frac{1}{2}(t_1 + t_2)$ equals t_n, or when the mean of the temperatures of the hot and cold junctions equals a certain temperature, constant for each pair of metals, there will be also no electromotive force in the circuit. This temperature t_n is that which has already been called the neutral temperature. The formula also assigns the value to that temperature t_1 at which, for fixed values of t_n and t_2, the electromotive force in the circuit is a maximum. If we represent the difference between t_n and t_1 by x, then $t_1 = t_n \pm x$. Using this value in the formula, we obtain $E_1 - E_2 = \frac{a}{2}((t_n - t_2)^2 - x^2)$. This is manifestly a maximum when $x = 0$. The electromotive force in a circuit is then, according to the formula, a maximum when the temperature of one junction is the neutral temperature.

The formula also shows that the thermoelectric power is zero when $t_1 = t_n$. We may set $E_1 = A + at_nt_1 - \frac{a}{2}t_1^2$. Now if t_1 take any small increment $\varDelta t_1$, E_1 has a corresponding increment $\varDelta E_1$. Hence we have $E_1 + \varDelta E_1 = A + at_nt_1 - \frac{a}{2}t_1^2 + at_n\varDelta t_1 - at_1\varDelta t_1$, if we neglect the term containing $\varDelta t_1^2$. From this equation we obtain $\frac{\varDelta E_1}{\varDelta t_1} = at_n - at_1$, which in the limit, as $\varDelta t_1$ becomes indefinitely small, is the thermoelectric power at the temperature t_1. It is positive for values of t_1 below t_n; is zero for $t_1 = t_n$, and negative for higher values of t_1. That is, if we assume $t_1 = t_2$ lower than t_n, and then gradually raise the temperature t_1, the thermoelectric power at the heated junction is at first positive, but

continually decreases in numerical value, until at $t_1 = t_n$ it becomes zero. At that temperature, then, the metals are thermoelectrically neutral to one another, and a small change in the temperature does not change the electromotive force at the junction.

317. The Thomson Effect.—Thomson has shown that, in certain metals, there must be a reversible thermal effect when the current passes between two unequally heated parts of the same metal. Let us suppose a circuit of copper and iron, of which one junction is at the neutral temperature and the other below the neutral temperature. The current then sets from copper to iron across the hot junction. In the hot junction there is no thermal effect produced, because the metals are at the neutral temperature. Across the cold junction the current is flowing from iron to copper, and hence is evolving heat. The current in the circuit can be made to do work, and since no other energy is imparted to the circuit this work must be done at the expense of the heat in the circuit. Since heat is not absorbed at either junction, it must be absorbed in the unequally heated parts of the circuit between the junctions.

To show this, Thomson used a conductor the ends of which were kept at constant temperatures in two coolers, while the central portion was heated. When a current was passed through this conductor, thermometers, placed in contact with exposed portions of the conductor between the heater and the coolers, indicated a rise of temperature different according as the current was passing from hot to cold or from cold to hot. The heat seems therefore to be carried along by the current, and the process has accordingly been called the *electrical convection of heat.* In copper the heat moves with the current, in iron against it. In another form of statement it may be said that, in unequally heated copper, a current from hot to cold heats the metal, and from cold to hot cools it, while in iron the reverse thermal effects occur. The experiments of Le Roux show that the process of electrical convection of heat cannot be detected in lead. For this reason lead is used as the standard metal in constructing the thermoelectric diagram.

CHAPTER VII.

LUMINOUS EFFECTS OF THE CURRENT

318. The Electric Arc.—If the terminals of an electric circuit, in which is an electromotive force of forty or more volts, be formed of carbon rods, a brilliant and permanent luminous arc will appear between the ends of the rods if they be touched together and then withdrawn a short distance from each other. The temperature of the arc is so high that the most refractory substances melt or are dissipated when placed in it. The carbon forming the positive terminal is hotter than the other. Both the carbons are gradually oxidized, the loss of the positive terminal being about twice as great as that of the negative. The arc is, however, not due to combustion, since it can be formed in a vacuum.

The current passing in the arc is, in ordinary cases, not greater than ten amperes, while the measurements of the resistance of the arc show that it is altogether too small to account for this current when the original electromotive force is taken into account. This fact has been explained by Edlund and others on the hypothesis that there is a counter electromotive force set up in the arc, which diminishes the effective electromotive force of the circuit. The measurements of Lang show that this counter electromotive force in an arc formed between carbon points is about thirty-six volts, and in one formed between metal points about twenty-three volts.

319. The Spark, Brush, and Glow Discharges.—When a conductor is charged to a high potential and brought near another conductor which is joined to ground, a *spark* or a series of sparks

will pass from one to the other. This phenomenon and others associated with it are most readily studied by the use of an electrical machine or an induction coil, between the electrodes of which a great difference of potential can be easily produced. If the spark be examined with the spectroscope, its spectrum is found to be characterized by lines which are due to the metals composing the electrodes, and to the medium between them.

The passage of the spark through air or any dielectric is attended with a sharp report, and if the dielectric be solid, it is perforated or ruptured. If the electrodes be separated by a considerable distance, the path of the spark is usually a zigzag one. It is probable that this is due to irregularities in the dielectric, due to the presence of dust particles.

With proper adjustment of the electrodes, the discharge may sometimes be made to take the form of a long *brush* springing from the positive electrode, with a single trunk which branches and becomes invisible before reaching the negative electrode. Accompanying this is usually a number of small and irregular brushes starting from the negative electrode.

Another form of discharge consists of a pale luminous *glow* covering part of the surface of one or both electrodes. If a small conducting body be interposed between the electrodes when the glow is established, a portion of the glow will be cut off, marking out a region on the electrode which is the projection of the intervening conductor by the lines of electrical force. This phenomenon is called the *electrical shadow*.

The difference of potential required to set up a spark between two slightly convex metallic surfaces, separated by a stratum of air 0.125 centimetre thick, has been shown by Thomson to be about 5500 volts. The difference of potential which produces the sparks between the electrodes of an electrical machine, which are sometimes fifty or sixty centimetres long, must therefore be very great. The quantity of electricity which passes during the discharge is, however, exceedingly small, on account of the great resistance of the medium through which the discharge takes place.

Faraday showed that many of the phenomena of the discharge depend to some extent upon the medium in which it occurs. The differences in color and in the facility with which various forms of the discharge were set up in the gases upon which he experimented were especially noticeable.

It was proved by Franklin that the lightning flash is an electrical discharge between a cloud and the earth or another cloud at a different electrical potential. The differences of potential to which such discharges are due must be enormous, and the heat developed by the discharge shows that the quantity of electricity which passes in it is considerable.

Slowly moving fire-balls are sometimes seen, which last for a considerable time and disappear with a loud report and with all the attendant phenomena of a lightning discharge. It is probable that they are glow discharges which appear just before the difference of potential between the cloud and the earth becomes sufficiently great to give rise to a lightning flash.

320. The Electrical Discharge in Rarefied Gases.—If the air between the electrodes of an electrical machine be heated, it is found that the discharge takes place with greater facility and that the spark which can be obtained is longer than before. Similar phenomena appear if the air about the electrodes be rarefied by means of an air-pump. After the rarefaction has reached a certain point the discharge ceases to pass as a spark, and becomes apparently continuous. The arrangement in which this discharge is studied consists of a glass tube into which are sealed two platinum or, preferably, aluminium wires to serve as electrodes, and from which the air is removed to any required degree of exhaustion by an air-pump. Such an arrangement is usually called a *vacuum-tube.*

As the exhaustion proceeds there appears about the negative electrode in the tube a bright glow, separated from the electrode by a small non-luminous region. The body of the tube is filled with a faint rosy light, which in many cases breaks up into a succession of bright and dark layers transverse to the direction of the discharge. The discharge in this case is called the *stratified discharge.* A

vacuum-tube in which the exhaustion is such that the phenomena are those here described is often called a *Geissler tube.* As the exhaustion is raised still higher, the rosy light in the tube fades out, the non-luminous space around the negative electrode becomes very much greater, and the phenomena in the tube become exceedingly interesting. They were discovered and have been carefully studied by Crookes, and the vacuum-tubes in which they appear are hence called *Crookes' tubes.* They may be most conveniently described by assuming that the molecules of gas in the tube break into their constituent ions in the region near the negative electrode, and that the negative ions are repelled from that electrode. The stream of negative ions may be called the *cathode discharge.* This view receives some support from the fact that the relations of current and resistance in the tube are such as to indicate a counter electromotive force at the negative electrode.

The region occupied by the discharge from the negative electrode may be recognized by a faint blue light, which was not visible in the former condition of the tube. At every point on the wall of the tube to which this discharge extends occurs a brilliant phosphorescent glow, the color of which depends on the nature of the glass. The discharge seems to be independent of the position of the positive electrode, and to take place in nearly straight lines, which start normally from the negative electrode. If two negative electrodes be fixed in the tube, the discharge from one seems to be deflected by the other, and two discharges which meet at right angles seem to deflect one another.

If the discharge from a flat electrode be made to fall upon a body which can be moved, such as a glass film, or the vane of a light wheel, mechanical motions will be set up.

If the negative electrode be made in the form of a spherical cup, and a strip of platinum-foil be placed at its centre, the foil will become heated to redness when the discharge is set up.

There is no evidence that two discharges in the same direction act directly on each other, but a magnet brought near the outside

of the tube will deflect a discharge as if it were an electrical current.

The explanation of these phenomena was indicated by Crookes, and Spottiswoode and Moulton. The particular form of it here given was developed by J. J. Thomson. It is assumed that they are due to the presence of the gas left in the tube after the exhaustion has been brought to an end. The mean free path of the molecules in the tube is much greater than that at ordinary densities, and they can accordingly move through long distances in the tube before their motion is checked by collisions. It is assumed that the molecules of gas in the tube are dissociated near the negative electrode, and that their negative ions are repelled from it. The phenomena which have been described are then due to the collision of these ions with other bodies or with the wall of the tube, or to their mutual electrical repulsions and to the action between a moving quantity of electricity and a magnet.

The experiments of Spottiswoode and Moulton, who showed that the same phenomena appeared at lower exhaustions, if the intensity of the discharge were increased, are in favor of this explanation. So is also the fact that the Crookes phenomena appear with a maximum intensity at a certain period during the exhaustion of the tube, while if the exhaustion be carried as far as possible, by the help of chemical means, they cease altogether and no current passes in the tube. The connection of these phenomena with the action of the radiometer (§ 222) is also at once apparent.

321. The Röntgen Radiance.—It was discovered by Hertz that the cathode discharge will pass through a thin strip of aluminium-foil placed in its path within the tube. In 1894 Lenard constructed a tube in which a part of the glass wall was replaced by aluminium-foil, and found that when the cathode discharge was directed upon the aluminium-foil a series of phenomena was obtained outside the tube, which he ascribed to the cathode discharge which passed through the aluminium. He found that similar effects could be produced outside a tube in which there was no aluminium window, and so concluded that the cathode discharge could pass through

glass; he also showed that it could pass through other substances with varying degrees of facility. Among the effects ascribed by Lenard to this discharge were the production of fluorescence in many fluorescent substances, the production of photographic action in ordinary photographic dry-plates, and the penetration of the discharge through bodies by an amount dependent upon their densities, it being less as the densities are greater. The discharge was deflected when brought into a magnetic field.

In 1895 Röntgen discovered that effects in some degree similar to those investigated by Lenard could be obtained from any highly exhausted vacuum tube. The results of his researches and of those of many other physicists who have investigated the same action may be described as follows: Wherever the cathode discharge falls upon certain substances, the most important of which, as yet known, are platinum and glass, an action is set up known as the *Röntgen radiance.* This radiance excites fluorescence in many fluorescent substances and acts upon the photographic plate. It proceeds in straight lines and its intensity varies inversely with the square of the distance; it is not affected by the presence of a magnetic field; in these respects it apparently differs from the action investigated by Lenard. It penetrates all substances and is partly obstructed by all substances, the obstruction being greater as the density of the substance is greater. It is apparently capable of true reflection to a very small degree. No indubitable evidence has as yet been given that it can be refracted, or that it exhibits the phenomena of interference, diffraction, or polarization. When it falls upon an electrified body the charge on the body gradually disappears, the effect being to render the air or other gas surrounding the body a conductor.

No satisfactory theory of the Röntgen radiance can as yet be given. It has been variously ascribed to the mechanical movement of the molecules of the residual gas in the tube in which it originates or of the walls of the tube, to transverse vibrations in the ether of a wave length much shorter than those of the shortest waves of light hitherto known, and to longitudinal vibrations in the ether.

The first explanation is supported by certain facts known with regard to the changes that go on in the tube as the discharge is kept up through it, but it is otherwise unsatisfactory. Most of the facts known are consistent with the theory of short transverse vibrations, but no explanation of their origin is given. The theory of longitudinal vibrations has been to some extent developed by Jaumann; he assumes that the characteristic factor of the dielectric, which we have called the dielectric constant, is not really constant, but variable, and a function of the electromotive force. He then shows that on this assumption the electrical discharge in rarefied gases may set up longitudinal waves, and that these waves possess many, if not all, of the properties of the cathode discharge. Since the properties of the cathode discharge and of the Röntgen radiance are not the same, we cannot conclude that the latter are explicable by longitudinal waves, though there is as yet no evidence to the contrary.

LIGHT.

CHAPTER I.

PROPAGATION OF LIGHT.

322. Vision and Light.—The ancient philosophers before Aristotle believed that vision consisted in the contact of some subtle emanation from the eye with the object seen. Aristotle showed the absurdity of this view by suggesting that if it were true, one should be able to see in the dark. Since his time it has been generally admitted that vision results from something proceeding from the body seen to the eye, and there impressing the optic nerve. This we call *light*.

Optics treats of the phenomena of light. It is conveniently divided into two branches: *Physical Optics*, which treats of the phenomena resulting from the propagation of light through space and through different media; and *Physiological Optics*, which treats of the sense of vision.

323. Theories of Light. The Ether. — The principal facts known about light in Newton's time, especially its propagation in straight lines, its reflection and refraction, could be explained by the hypothesis that light consisted of small material particles or corpuscles emitted from luminous bodies with very high velocities. This *emission theory* was adopted and defended by Newton.

Newton's contemporary Huygens proposed to explain the phenomena of light as the result of waves set up in luminous bodies and transmitted by an elastic medium which pervades all space. The properties which Huygens assigned to this medium were those

of a fluid, in which only longitudinal waves, like those of sound, can exist, and they were therefore inadequate to explain polarization, which shows that light in some way differs in different directions in the wave front, or, as Newton expressed it, has sides; and further, since observation and theory, so far as it was then developed, showed that a wave which has passed through an opening will spread in all directions from that opening and will not be propagated in a line, as light is, Newton rejected the wave theory, and developed the emission theory by assigning such properties to the light corpuscles as were needed to explain the facts then known.

The discovery by Young of interference, which can be most easily explained by the wave theory, and the demonstration by Fresnel that rectilinear propagation can be explained by taking into account the shortness of the wave length, and the further bold assumption by Fresnel that the properties of the medium are those of a solid and that the vibrations in the light wave are transverse to the line of progress, removed the objections which had been felt by Newton and set the *wave theory* on a secure foundation. The one objection which was still felt arose from the necessity of assuming the existence of an all-pervading medium, which cannot be made evident to any of our senses and which possesses properties unlike those of any known body. This objection has gradually disappeared in view of the almost complete success attained by the wave theory in explaining the phenomena of light. The demonstration by Maxwell that all magnetic and electrical phenomena can be explained by actions in a similar medium, and that the properties of electromagnetic waves in such a medinm are precisely the same as those of light waves, has done much to strengthen the evidence for the existence of this medium. Its properties are probably not those of an elastic solid, but they are such as to enable us to represent light waves as if they were waves in an elastic solid; we will accordingly, in what follows, use this mode of representation, with the understanding that the rigidity and density which are ascribed to the medium are representative of other properties of the medium, which, in respect to light waves, are equivalent properties.

The medium in which magnetic, electrical, and light phenomena take place is called the *ether*. It pervades all space within the bounds of the known universe, and is so far material that it can transmit energy from one material body to another; the manner of its connection with the atoms of matter is not well understood, and the question of the influence of matter upon it is one of the most obscure in modern physics. The ether was first represented by Fresnel as an elastic solid possessing a rigidity estimated by Thomson to be about the one-thousand millionth of that of steel, and a density estimated to be 9.36×10^{-19} grams per cubic centimetre. Thomson has shown that the properties of the ether, at least those concerned in the transmission of light, may be explained by supposing it composed of minute material bodies rotating like gyroscopes about definite axes. The most interesting view of the ether is that recently proposed by Fitzgerald. He conceives of it as a continuous fluid filled with vortices. These vortices may be either infinitely long linear vortices threading past each other in all directions, or ring vortices interlinked with each other; Fitzgerald has shown that such an assemblage of vortices will transmit electromagnetic vibrations comparable in all respects to those of light. The connection of this theory with Thomson's theory of the vortex atom gives it additional interest.

324. Wave Surfaces.—In § 130 is explained the general mode of propagation of wave motion in accordance with Huygens' principle. When light emanating from a point proceeds with the same velocity in all directions, the wave fronts are evidently concentric spherical surfaces. There are, however, many cases, especially in crystalline bodies, of unequal velocities in different directions. In these cases the wave fronts are not spherical, but ellipsoidal, or surfaces of still greater complexity.

325. Straight Lines of Light.—When a small screen *A* (Fig. 92) is placed between the eye and a luminous point, the luminous point is no longer visible. Light cannot reach the eye by the curved or broken line *PAE*, and is therefore said to move in straight lines. This seems not to accord with Huygens' principle,

which makes any wave front the resultant of an infinite number of elementary waves proceeding from the various points of the same

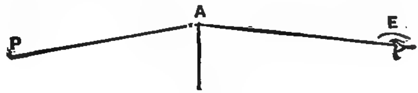

FIG. 92.

wave front in one of its earlier positions. It can, however, easily be shown that when the wave lengths are small, the disturbance at any point P (Fig. 93) is due almost wholly to a very small portion of the approaching wave. Let us consider first the case of an isotropic medium, in which light moves in all directions with the same velocity. Let mn be the front of a linear wave perpendicular to the plane of the paper, moving from left to right or towards P. Draw PA perpendicular to the wave front, and draw Pa, Pb, etc., at such obliquities that Pa shall exceed PA by half a wave length, Pb exceed Pa by half a wave length, etc. We will designate the wave length by λ.

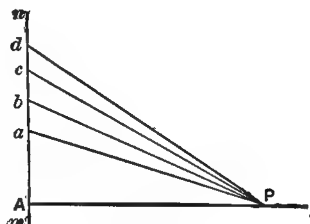

FIG. 93.

It is evident that the total effect at P will be the sum of the effects due to the small portions Aa, ab, etc., called half-period elements. Since Pa is half a wave length greater than PA, and Pb half a wave length greater than Pa, each point of ab is half a wave length farther from P than some point in Aa; hence elementary waves from ab will meet at P waves from Aa in the opposite phase. It appears, therefore, that the effects at P of the portions ab and Aa are opposite in sign, and tend to annul each other. The same is true of bc and cd. But the effects of Aa and ab may be considered as proportional to their lengths. Hence, by computing the lengths, we can determine the resultant effect at P. Let $AP = x$. From the construction we have

$$Aa = \sqrt{\left(x+\frac{\lambda}{2}\right)^2 - x^2} = \sqrt{x\lambda + \frac{\lambda^2}{4}};$$

$$Ab = \sqrt{(x+\lambda)^2 - x^2} = \sqrt{2x\lambda + \lambda^2};$$

$$Ac = \sqrt{(x+\tfrac{3}{2}\lambda)^2 - x^2} = \sqrt{3x\lambda + \tfrac{9}{4}\lambda^2};$$

$$Ad = \sqrt{(x+2\lambda)^2 - x^2} = \sqrt{4x\lambda + 4\lambda^2};$$

$$\text{etc.} = \text{etc.}$$

For light the values of λ are between 0.00039 and 0.00076 mm., and if x be taken as 1000 mm., λ^2 will be very small in comparison with $x\lambda$, and may be omitted. The above formulas then become, if $\sqrt{x\lambda}$ be represented by l,

$$Aa = l\sqrt{1};\ Ab = l\sqrt{2};\ Ac = l\sqrt{3};\ Ad = l\sqrt{4};\ \text{etc.} = \text{etc.},$$

and the several portions into which the wave front is divided are

$$Aa = l = 1l; \qquad ab = l(\sqrt{2}-1) = 0.414l;$$

$$bc = l(\sqrt{3}-\sqrt{2}) = 0.318l; \quad cd = l(\sqrt{4}-\sqrt{3}) = 0.268l.$$

Taking now the pairs of which the effects at P are opposite in sign, we find Aa a little more than twice ab, while bc and cd are nearly equal. It is evident, also, that for portions beyond d adjacent pairs will be still more nearly equal, and the effect at P, therefore, of each pair of segments beyond b almost vanishes. The effect at P is then almost wholly due to that portion of Aa that is not neutralized by ab. But, taking the greatest value of λ, $Aa = \sqrt{x\lambda} = \sqrt{0.76} = 0.87$ mm., a very small distance. Hence, under the conditions assumed, the effect at any point P is due to that portion of the wave-front near the foot of the perpendicular let fall from P on the wave-front. It may be demonstrated by experiment that the portions of the wave beyond Aa neutralize each other. Suppose a screen mn in the position shown in Fig. 94. The point P will be in shadow. If the darkness at P is due

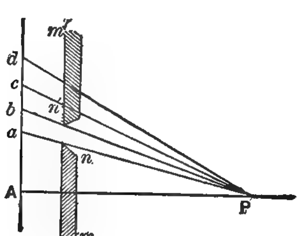

FIG. 94.

to interference as explained, light should be restored by suppressing the interfering waves. If a second screen be placed at $m'n'$ so as to cut off the waves proceeding from points above b, waves from points between a and b will no longer be neutralized, and light should fall at P. To test this conclusion the edge of a flat flame may be observed through a narrow slit in a screen. Instead of the narrow edge of the flame, a broad luminous surface is seen, in which the brightness gradually diminishes from the centre towards the edges. If we consider the wave-front just entering the slit, it will be seen that elementary waves proceed from all points of it, and the slit being very narrow it is only in very oblique directions that pairs of these waves can meet in opposite phases. Hence light proceeds in oblique lines behind the screen, and from our habit of locating visible objects back along the line of light entering the eye, the flame appears as a broad surface. It will be seen by reference to Fig. 93 that the elementary wave that first reaches P is the one to which the disturbance there is principally due. Other waves arriving later find there the opposite phase of some wave that has preceded them. When the velocity in all directions is the same, the first wave to reach P is the one that starts from the foot of a perpendicular let fall from P on the wave-front. Hence light is said to travel in straight lines perpendicular to the wave-front. If, however, light does not move with equal velocities in all directions, the last statement is no longer true, as will be seen from Fig. 95. Here mn represents a wave-front, proceeding towards P in a medium in which the velocities in different directions are such that the elementary wave-surfaces are ellipsoids. The ellipses in the figure may be taken as sections of these ellipsoids. The wave first to reach P is not the one that starts from A at the foot of the perpendicular, but from A'. It is from A' that P derives its light, and the line of propagation is no longer perpendicular to the wave-front.

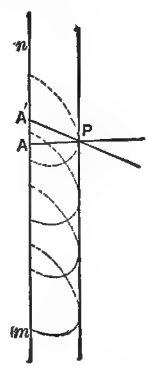

FIG. 95.

The demonstration here given fails when applied to a plane

wave, and some other explanation must be given for the rectilinear propagation of light in such a wave. For consider a plane wave advancing toward a point P, and describe on it a series of circles, the distances of which from P differ by half a wave-length. These circles cut a line in the surface drawn from A (Fig. 93), the foot of the perpendicular to the wave-front from P, in the points a, b, c, etc., and the rings enclosed between them are half-period elements. The areas of these rings are $2\pi\overline{Aa}^2$, $2\pi(\overline{Ab}^2 - \overline{Aa}^2)$, $2\pi(\overline{Ac}^2 - \overline{Ab}^2)$, etc. If λ be very small, they each become equal to $2\pi l^2$. Hence if their effects at P depend only on their areas, they would annul one another, and no light would reach P. We are therefore forced to assume that the effect of each area in sending light to P diminishes as the obliquity increases, so that the first area is more efficient than the second, the second than the third, and so on. The effectiveness of the areas diminishes at first slowly, and afterwards more rapidly, the more distant areas having nearly the same efficiency. Representing the efficiency of the areas by m_1, m_2, m_3, etc., and remembering that the even areas oppose the action of the odd ones, we may write the total efficiency in the form $\frac{1}{2}m_1 + \frac{1}{2}(m_1 - m_2) - \frac{1}{2}(m_2 - m_3) + \frac{1}{2}(m_3 - m_4) - \ldots$. Each of the terms in parenthesis is very nearly equal to zero, and the efficiency at P is therefore nearly half of that of the central area. The light therefore appears to reach P from a small area around A.

It is important to note that the deductions of this section apply only where λ is small in relation to x, so that λ^2 may be neglected in comparison with $x\lambda$. With sound-waves this is not true, and if a computation similar to that given above for light-waves be made for sound, not omitting λ^2, it will be seen why there are no definite straight lines of sound and no sharp acoustic shadows.

326. Principle of Least Time.—The above are only particular cases of a law of very general application, that light in going from one point to another follows the path that requires least time. The reason is that values in the vicinity of a minimum change slowly, and there will be a number of points in the neighborhood of that point from which the light-waves are propagated to the given point

in the least time, from which waves will proceed to that point in sensibly the same time, and, meeting in the same phase, combine to produce light. It is also true that values change slowly in the vicinity of a maximum, and there are cases where the path followed by the light is determined by the fact that the time is a maximum instead of a minimum.

327. Shadows.—An optical *shadow* is the space from which light is excluded by an opaque body. When the luminous source is a point, or very small, the boundary between the light and shadow is very sharp. When the luminous source is large, there is a portion of the space behind the opaque body, called the *umbra*, which is in deep shadow, and surrounding this is a space which is in shadow with reference to one portion of the luminous source while it is in the light with reference to another portion. The space from which light is only partially excluded is the *penumbra*. Fig. 96 shows the

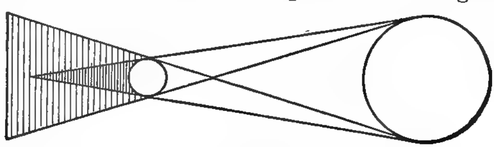

FIG. 96.

boundaries of the umbra and penumbra. It is evident that the light diminishes gradually from the outer boundary of the penumbra to the boundary of the umbra.

328. Images by Small Apertures.—If light from a single luminous point pass through a small hole of any form, and fall on a screen at some distance, it produces a luminous spot of the same form as the opening. Light from several points will produce several such spots. If the luminous source be a surface, the spots produced by the light from its several points will overlap each other and form an illuminated surface, which, if the source be large in comparison with the opening, will have the general form of the source, and will be inverted. The illuminated surface is an inverted *image* of the source. If a small opening be made in the window-shutter of a darkened room, images of external objects will be seen on the wall opposite. The smaller the opening, the more sharply defined, but the less brilliant, is the image.

VELOCITY OF LIGHT.

329. Velocity Determined from Eclipses of Jupiter's Moons.—Roemer, a Danish astronomer, was led to assume a progressive motion for light in order to explain some apparent irregularities in the motions of Jupiter's satellites. A few observations of one of Jupiter's moons are sufficient to determine the time of its eclipses for months in advance. If these observations be made when the earth and Jupiter are on the same side of the sun, and the time of an eclipse occurring about six months later predicted from them be compared with the observed time of that eclipse, it is found that the observed time is about $16\frac{2}{3}$ minutes later than the predicted time. This discrepancy is explained if it be assumed that light has a progressive motion and requires $16\frac{2}{3}$ minutes to cross the earth's orbit, for the distance of the earth from Jupiter in the second case is about the diameter of its orbit greater than in the first.

330. Aberration of the Fixed Stars.—The apparent direction of the light coming from a star to the earth, that is, the apparent direction of the star from the earth, is the resultant of the motion of the light and the motion of the earth. As the motion of the earth changes direction the apparent direction of the star will change also, and the amount of that change will depend on the relation between the velocity of light and the change in the velocity of the earth in its orbit, understanding by change of velocity change in direction as well as in amount. This apparent change in the position of the stars is called *aberration*. Knowing its amount corresponding to a known change in the earth's motion, we may compute the velocity of light. This method was first employed by Bradley.

Though the agreement of the velocity of light thus determined with that measured by other methods seems to confirm the validity of the reasoning here given, yet there are serious and unexplained difficulties in the theory of aberration, arising from the discrepant results of experiments instituted to determine the relations of the ether to bodies moving through it.

331. Fizeau's Method.—Several methods have been employed for measuring the velocity of light by determining the time required

for it to pass over a small distance on the earth's surface. In the form of experiment devised by Fizeau, a beam of light is allowed to pass out through a small hole in the shutter of a darkened room to a distant station, where it is reflected back on itself. It returns through the opening and produces an image of the source. A toothed wheel is placed in front of the opening in such a position that, to pass out or back, the light must pass through the spaces between the teeth. If the wheel revolve slowly, as each space passes the opening in the shutter light will pass out, and returning from the distant station will enter through the space by which it made its exit. An image of the source will therefore be visible whenever a space passes the opening, and in consequence of the persistence of vision this image will appear continuous. Since it takes time for the light to go to the distant station and back, it is possible to give to the wheel such a velocity that when the light which passed out through a given space returns, it will find the adjacent tooth covering the opening, so that no image of the source can be seen. If the velocity of rotation be sufficiently increased, the image again comes into view when the light can enter through the space following that by which it emerged. A still further increase of velocity may cause a second extinction of the image. The experiment consists in determining accurately the velocities for which the several extinctions and reappearances of the image occur. A high degree of accuracy cannot be attained because the extinction of the image is not sudden. It disappears by a gradual fading away, and reappears by a gradual brightening. For quite a range of velocity the image cannot be seen at all.

332. Foucault's Method.—Foucault's method depends upon the use of the revolving mirror as a means of measuring a very small interval of time. Foucault's experiments were repeated with some modification by Michelson in 1879 and again in 1882. The general theory of the experiment may be understood from the following brief description: Let S (Fig. 97) be a narrow slit, m a mirror which may revolve about an axis in its own plane, L a lens, and m' a second mirror. Light from a source behind S passes through the slit,

falls on m, is reflected, when m is in a suitable position, through the lens L, and forms an image at S'. S and S' are conjugate foci of

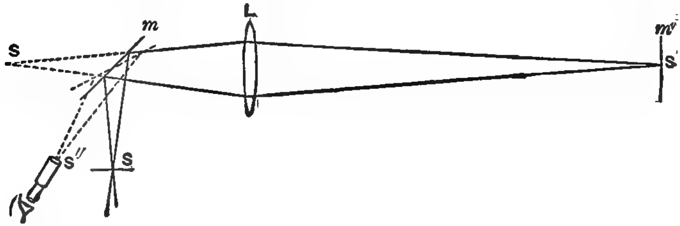

FIG. 97.

the lens, and by so placing the lens that S shall be a little beyond the principal focus, S' may be removed to as great a distance as desired. The mirror m' is perpendicular to the axis of the lens, and at such a distance that the image S' falls upon its surface. It is evident that any light reflected back from m' through L will return to the conjugate focus S, whatever the position of the mirror m', so long as it sends the light in such a direction as to pass through L both going and returning. If now the mirror m be given a rapid rotation clockwise, light passing through L will return to find m in a changed position, and the image will be displaced from S to some point S'' to the left of S. Knowing the displacement SS'' and the number of rotations of the mirror per second, the time required for light to pass from m to S' and back is determined. The value of the velocity of light, as determined by Michelson in 1879, is 299,910, and in 1882, 299,853, kilometres per second.

CHAPTER II.

REFLECTION AND REFRACTION.

333. Law of Reflection.—In § 132 it is shown that when a wave passes from one medium into another where the particles constituting the wave move with greater or less facility, a wave is propagated back into the first medium. It is shown in § 133, that when the surface separating the two media is a plane surface, the centres of the incident and reflected waves are on the same perpendicular to the surface, and at equal distances on opposite sides. Considering the lines to which, as shown in § 325, the wave propagation in the case of light is restricted, a very simple law follows at once from this relation of the incident and reflected waves. In Fig. 98, C and C' represent the centres of the incident and reflected waves *mn*, *on*. CA, AB are the paths of the incident and reflected light. It will be evident from the figure that CA, AB are in the same plane normal to the reflecting surface, and that they make equal angles with the normal AN. CAN is called the angle of incidence, and NAB the angle of reflection. Hence we may state the *law of reflection* as follows: The angles of incidence and reflection are equal, and lie in the same plane normal to the reflecting surface. By constructing half-period elements in the reflecting surface, it can easily be shown that the portion of the wave from which light reaches B is that lying around A. This may likewise be shown by proving, as can easily be done, that light traverses the path CAB from C to B which fulfils this law, in less time than it requires to traverse any other path by way of the reflecting surface.

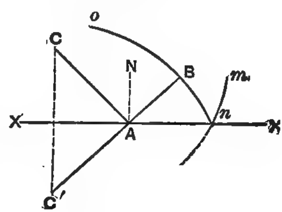

FIG. 98.

334. Law of Refraction.—If the incident wave pass from the one medium into the other, there is, in general, a change in the wave front, and a consequent change in the direction of the light. Let us first consider the simple case of a plane wave entering a homogeneous, isotropic medium of which the bounding surface is plane. Suppose both planes perpendicular to the plane of the paper, and let AB (Fig. 99) represent the intersection of the surface of the medium, and mn the intersection of the wave with that

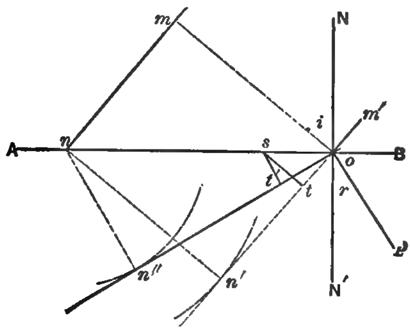

FIG. 99.

plane. Let v represent the velocity of light in the medium above AB, and v' the velocity in the medium below it. Let $m'o$ be the position of the wave in the first medium after a time t. Then mo equals vt. As the wave front passes from mn to $m'o$, the points of the separating surface between n and o are successively disturbed, and become centres of spherical waves propagated into the second medium with the velocity v'. The wave surface of which the centre is n would, at the end of time t, have a radius $nn'' = v't$, such that $\frac{nn'}{nn''} = \frac{v}{v'}$. Similarly, the wave from any other point, as s, would have a radius st' such that $\frac{st}{st'} = \frac{v}{v'}$, and the wave surface within the second medium is evidently the plane on''. As the direction of propagation is perpendicular to the wave front, op will represent the direction of the light in the second medium. In the triangles non' and non'' we have $nn' = no \sin Aon'$, and

$$nn'' = no \sin Aon''; \quad \text{hence} \quad \frac{\sin Aon'}{\sin Aon''} = \frac{nn'}{nn''} = \frac{v}{v'}.$$

If we represent the *angle of incidence moN* by i, and the *angle of refraction poN'* by r, we have

$$\frac{\sin i}{\sin r} = \frac{v}{v'} = \mu, \text{ a constant.} \tag{108}$$

This constant is called the *index of refraction.* This is the expression of Snell's *law of refraction.* Here again the time required for the light to pass by *mop* from m in one medium to p in the other is less than by any other path.

We may now trace a wave through a medium bounded by plane surfaces. Suppose the wave front and bounding planes of the medium all perpendicular to the plane of the paper. We shall have as above for the first surface $\frac{\sin i}{\sin r} = \frac{v}{v'} = \mu$, and for the second surface $\frac{\sin i'}{\sin r'} = \frac{v'}{v''} = \mu'$.

If, as is often the case, the light emerge into the first medium, $v'' = v$, and $\frac{\sin i'}{\sin r'} = \frac{v'}{v} = \frac{1}{\mu}$.

If the bounding planes be parallel, $i' = r$, and we have $\frac{\sin r}{\sin r'} = \frac{1}{\mu}$; hence $i = r'$, or the incident and emergent waves are parallel. If the two bounding planes form an angle A the body is called a *prism.* The wave incident upon the second face will make with it an angle $A - r$, and the emergent wave is found by the relation $\frac{\sin (A-r)}{\sin r'} = \frac{1}{\mu}$ or $\frac{\sin r'}{\sin (A-r)} = \mu$. The direction of the emerging wave front may be found by construction.

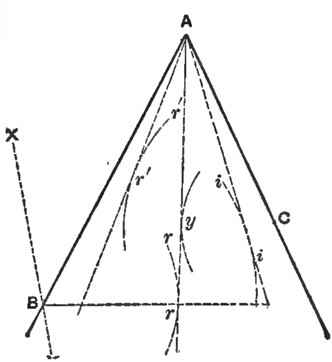

FIG. 100.

Draw Ai (Fig. 100) parallel to the incident wave. From some point B on AB describe an arc tangent to Ai; from the same point with a radius $\frac{Bi}{\mu}$ describe the arc rr. Ar, tangent to rr, is the refracted wave front. From some point C on

AC describe an arc tangent to Ar, and from the same point as centre describe another arc $r'r'$ with a radius $\mu \times Cy$. A tangent from A to $r'r'$ is parallel to the emergent wave. It might be that A would fall inside the arc $r'r'$ so that no tangent could be drawn. That would mean that there could be no emergent wave. The angle of incidence for which this occurs can readily be obtained. We have $\frac{\sin i'}{\sin r'} = \frac{1}{\mu}$, or $\sin r' = \mu \sin i'$. Now the maximum value of $\sin r'$ is 1, which is reached when $\sin i' = \frac{1}{\mu}$. Any larger value of $\sin i'$ gives an impossible value for $\sin r'$. The angle $i' = \sin^{-1} \frac{1}{\mu}$ is called the *critical angle* of the substance. For larger angles of incidence the light cannot emerge, but is *totally reflected* within the medium.

If a beam of white light be allowed to fall upon a prism through a narrow slit, it will be refracted, in general, in accordance with the law already given. The image of the slit, however, when projected upon a screen, appears not as a single line of white light, but as a variously colored band. This is due to the fact that the indices of refraction for light of different colors are different. Hence the index of refraction of a substance, as ordinarily given, depends upon the color of the light used in determining it, and has no definite meaning unless that color is stated.

335. Plane Mirrors.—The wave *on*, represented in Fig. 98, is the same as would have come from a luminous point at C' if the reflecting surface did not intervene. If this wave reach the eye of an observer, it has the same effect as though coming from such a point, and the observer apparently sees a luminous point at C'. C' is a *virtual image* of C. When an object is in front of a plane mirror each of its points has an image symmetrically situated in relation to the mirror, and these constitute an image of the object like the latter in all respects, except that by reason of symmetry it is reversed in one direction.

336. Spherical Mirrors.—A spherical mirror is a portion of a

spherical surface. It is a concave mirror if reflection occur on the concave or inner surface; a convex mirror if it occur on the convex surface. The centre of the sphere of which the mirror forms a part is its *centre of curvature.* The middle point of the surface of the mirror is the *vertex.* A line through the centre of curvature and the vertex is the *principal axis.* Any other line through the centre of curvature is a *secondary axis.* The angle between radii drawn to the edge of the mirror on opposite sides of the vertex is the *aperture.* To investigate the effects of reflection from a spherical surface, let us consider first a concave mirror. Let a light-wave emanate from a point L on the principal axis (Fig. 101). In general, different points of the wave

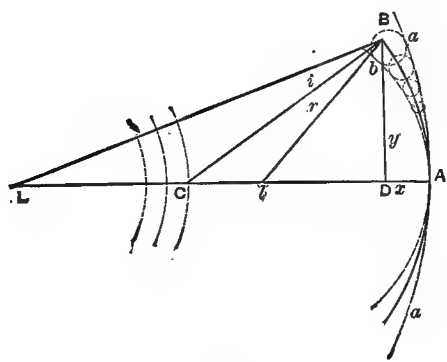

FIG. 101.

will reach the mirror successively, and, considering the elementary waves that proceed in turn from its several points, the reflected wave surface may be constructed as for a plane mirror. If the mirror were not there the wave front would, at a certain time, occupy the position *aa*. Drawing the elementary wave surfaces we have *bb*, the position at that instant of the reflected wave. Its form suggests that of a spherical surface, concave toward the front, and having a centre at some point on the axis. The elementary waves at B will certainly send light to some point l on the axis. We will examine the conditions which must be fulfilled in order that

all other points of the mirror shall send light to the same point, or that the reflected wave shall be a sphere with its centre at l. In order that this shall be the case, the distance $LB + Bl$ must be constant wherever the point B is situated on the reflecting surface. Draw BD perpendicular to the axis of the mirror. Represent BD by y, AD by x, LA by p, lA by p', and CA by r. Then we have $LB = \sqrt{(p - x)^2 + y^2}$, and $y^2 = (2r - x)x = 2rx - x^2$. Hence follows

$$LB = \sqrt{p^2 - 2px + x^2 + 2rx - x^2} = \sqrt{p^2 + 2x(r - p)}.$$

If the aperture be small, x will be small in comparison with the other quantities, and we may obtain the value of LB to a near approximation by extracting the root of this expression and omitting terms containing the second and higher powers of x. We obtain

$$LB = p + \frac{x}{p}(r - p) + \ldots.$$

In like manner we have

$$lB = p' + \frac{x}{p'}(r - p') + \ldots,$$

whence $$LB + lB = p + p' + \frac{x}{p}(r - p) + \frac{x}{p'}(r - p').$$

When B coincides with A, the above value becomes $p + p'$, and the condition that all values of $LB + lB$ are equal, whatever be the value of x within the limits already set to it, is found from

$$p + p' = p + p' + \frac{x}{p}(r - p) + \frac{x}{p'}(r - p').$$

From this equation we obtain $\frac{r}{p} + \frac{r}{p'} = 2$ and $p' = \frac{pr}{2p - r}$.

For the apertures for which the approximations by which the result was arrived at are admissible, the wave surface is practically spherical, and the point, the distance of which from the mirror is given by this equation, is the centre of the reflected wave. Since the disturbances propagated from bb reach l simultaneously, their effects are added, and the disturbance at l is far greater than at any other point. The effect of the wave motion is concentrated at l,

and this point is therefore called a *focus*. Since the light passes through l, it is a *real focus*. If l were the radiant point, it is clear that the reflected light would be concentrated at L. These two points are therefore called *conjugate foci*. If we divide both sides of the equation $\frac{r}{p}+\frac{r}{p'}=2$ by r, we have

$$\frac{1}{p}+\frac{1}{p'}=\frac{2}{r} \tag{109}$$

which is the usual form of the equation used to express the relation between the distances of the conjugate foci from the mirror.

A discussion of this equation leads to some interesting results. Suppose $p=\infty$, then $p'=\frac{1}{2}r$; that is, when the radiant is at an infinite distance from the mirror, the focus is midway between the mirror and the centre. In this case the incident wave is normal to the principal axis, and the focus is called the *principal focus*. Suppose $p=r$; $p'=r$ also. When $p=\frac{1}{2}r$, $p'=\infty$. When $p<\frac{r}{2}$, $\frac{1}{p}>\frac{2}{r}$ and $\frac{1}{p'}=\frac{2}{r}-\frac{1}{p}$, a negative quantity. To interpret this negative result it should be remembered that all the distances in the formulas were assumed positive when measured from the

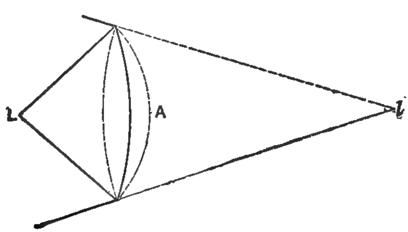

FIG. 102.

mirror toward the source of light. A negative result means that the distance must be measured in the opposite direction, or behind the mirror. Fig. 102 represents this case. It is evident that the reflected wave is convex toward the region it is approaching, and proceeds as though it had come from l. l is therefore a *virtual focus*. Either of the other quantities of the formula may have negative values. p will be negative if waves approaching their

centre l fall on the mirror. Plainly they would be reflected to L at a distance from the mirror less than $\frac{r}{2}$, as may be seen from the formula. If r be negative, the centre is behind the mirror. The mirror is then convex, and the formula shows that for all positive values of p, p' is negative and numerically smaller than p.

337. Refraction at Spherical Surfaces.—The method of discussion which has been applied to reflection may be employed to study refraction at spherical surfaces. Let BD (Fig. 103) be a spherical

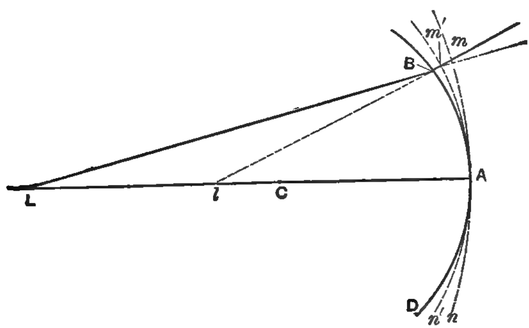

FIG. 103.

surface separating two transparent media. Let v represent the velocity of light in the first medium, to the left, and v' the velocity in the second medium, to the right, of BD. Let L be a radiant point, and mn a surface representing the position which the wave surface would have occupied at a given instant had there been no change in the medium, $m'n'$ the wave surface as it exists at the same instant in the second medium in consequence of the different velocity of light in it, and l the point where the prolongation of Bm' backwards cuts the axis. We will investigate the conditions which must be fulfilled in order that the refracted wave shall appear to proceed from a point on the axis, or shall be a spherical wave.

In order that this should be the case, the time occupied by the light in travelling from the point l on the axis with the velocity

which it has in the second medium should be the same for all points on the refracted wave. This time is given by $t' = \frac{lm'}{v'} = \frac{lB + Bm'}{v'}$; we are to investigate the conditions that this shall be the same for all points on the refracted wave.

The time occupied by the light in travelling from L to m' is $\frac{LB}{v} + \frac{Bm'}{v'} = C$, a constant for all points on the refracted wave. Subtracting from this the expression for t', we have $C - t' = \frac{LB}{v} - \frac{lB}{v'}$, and the condition that t' should be constant is therefore that $\frac{LB}{v} - \frac{lB}{v'}$ is constant. Since $\frac{v}{v'} = \mu$, we may write the expression which should be constant $LB - \mu lB = LA - \mu lA$.

Using the notation of the last section, and substituting the values of LB and lB as there found, except that p'' is used instead of p', we have as the condition that t' is a constant, whatever be the value of x within the limits set to it, the equation

$$p + \frac{x}{p}(r - p) - \mu\left(p'' + \frac{x}{p''}(r - p'')\right) = p - \mu p''.$$

From this we obtain $\frac{r}{p} - \frac{\mu r}{p''} = 1 - \mu$,

and
$$\frac{\mu}{p''} - \frac{1}{p} = \frac{\mu - 1}{r}. \qquad (110)$$

Hence the point at the distance p'' from the centre of the refracting surface is the centre of a spherical refracted wave.

If the medium to the right of BD be bounded by a second spherical surface, it constitutes a *lens*. Suppose this second surface to be concave toward l and to have its centre on AC. The wave $m'n'$, in passing out at this second surface, suffers a new change of form precisely analogous to that occurring at the first surface, and the new centre is given by the formula just deduced by substituting for p the distance of the wave centre from the new surface, and for μ the index of refraction of the third medium in relation to the second. If s represent the distance of l from the new

surface, μ' the new index, and p' the new focal distance, we have $\frac{\mu'}{p'} - \frac{1}{s} = \frac{\mu' - 1}{r'}$.

If we suppose the lens to be very thin, we may put $s = p''$. If we suppose also that the medium to the right is the same as that to the left of the lens, μ' is equal to $\frac{1}{\mu}$. On these suppositions $\frac{\frac{1}{\mu}}{p'} - \frac{1}{p''} = \frac{\frac{1}{\mu} - 1}{r'}$. Multiplying through by μ, we have

$$\frac{1}{p'} - \frac{\mu}{p''} = \frac{1-\mu}{r'} = -\frac{\mu - 1}{r'}.$$

Eliminating p'' between this equation and equation (110), we obtain

$$\frac{1}{p'} - \frac{1}{p} = (\mu - 1)\left(\frac{1}{r} - \frac{1}{r'}\right), \qquad (111)$$

which expresses the relation between the conjugate foci of the lens. It should be noted that r in the above formulas represents the radius of the surface on which the light is incident, and r' that of the surface from which the light emerges. All the quantities are positive when measured toward the source of light. Fig. 104

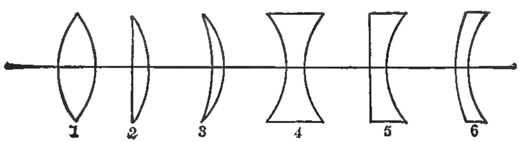

FIG. 104.

shows sections of the different forms of lenses produced by combinations of two spherical surfaces, or of one plane and one spherical surface.

An application of equation (111) will show that for the first three, which are thickest at the centre, light is concentrated, and for the second three diffused. The first three are therefore called *converging*, and the second three *diverging*, lenses. Let us consider the first and fourth forms as typical of the two classes. The first

is a double-convex lens. The r of equation (111) is negative because measured from the lens away from the source of light. The second term of the formula has therefore a negative value, and p' is negative except when $\frac{1}{p} > (\mu - 1)\left(\frac{1}{r} - \frac{1}{r'}\right)$. If $p = \infty$, we have $\frac{1}{p} = 0$ and $\frac{1}{p'} = (\mu - 1)\left(\frac{1}{r} - \frac{1}{r'}\right)$, a negative quantity because r is negative. p' is then the distance of the principal focus from the lens, and is called the *focal length* of the lens. The focal length is usually designated by the symbol f. Its negative value shows that the principal focus is on the side of the lens opposite the source of light. This focus is real, because the light passes through it. Equation (111) is a little more simple in application if, instead of making the algebraic signs of the quantities depend on the direction of measurement, they are made to depend on the form of the surfaces and the character of the foci. If we assume that radii are positive when the surfaces are convex, and that focal distances are positive when foci are real, the signs of p' and r in that equation must be changed, since in the investigation p' is the distance of a virtual focus, and r the radius of a concave surface. The formula then becomes

$$\frac{1}{p'} + \frac{1}{p} = (\mu - 1)\left(\frac{1}{r} + \frac{1}{r'}\right). \tag{112}$$

To apply this formula to a double-concave lens, r and r' are both negative; p' is then negative for all positive values of p. That is, concave lenses have only virtual foci. For a plano-convex lens (Fig. 104, 2), if light be incident on the plane surface, $r = \infty$ and $\frac{1}{p'} = (\mu - 1)\frac{1}{r'} + \frac{1}{p}$.

This gives positive values of p' and real foci for all values of $\frac{1}{p} < (\mu - 1)\frac{1}{r'}$.

For a concavo-convex lens (Fig. 104, 6) the second member of the equation will be negative, since the radius of the concave surface is negative and less numerically than that of the convex

surface. Hence p' is always negative and the focus virtual when L is real.

338. Images formed by Mirrors.—In Fig. 105 let ab represent an object in front of the concave mirror mn. We know from what precedes that if we consider only the light incident near c, the light reflected will be concentrated at some point a' on the axis ac at a distance from the mirror given by equation (109).

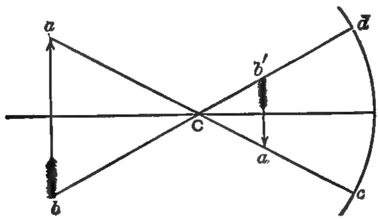

FIG. 105.

a' is a *real image* of a. In the same way b' is an image of b. If axes were drawn through other points of the object, the images of those points would be found in the same way. They would lie between a' and b', and $a'b'$ is therefore a real image of the object. It is inverted, and lies between the axes ac, bd, drawn through the extreme points of the object. The ratio of its size to that of the object is seen from the similar triangles abC, $a'b'C$, to be the ratio of the distances from C. From equation (109) we obtain $\frac{p'}{p} = \frac{r}{2p - r} = \frac{r - p'}{p - r}$.

Since $r - p'$ and $p - r$ are respectively the distances from the centre of the image and object, we have $\frac{a'b'}{ab} = \frac{r - p'}{p - r} = \frac{p'}{p}$; or, the image and object are to each other in the ratio of their respective distances from the mirror. As the object approaches, the image recedes from the mirror and increases in size. At the centre of curvature the image and object are equal, and when the object is within the centre and beyond the principal focus the image is outside the centre and larger than the object. When the object is between the principal focus and the mirror, the image is virtual

and larger than the object. Convex mirrors produce only virtual images, which are erect and smaller than the object.

339. Images formed by Lenses.—Let us suppose an object in front of a double-convex lens, which may be taken as a type of the converging lenses. The point c (Fig. 106) will have an image at the conjugate focus on the principal axis. a and b will have im-

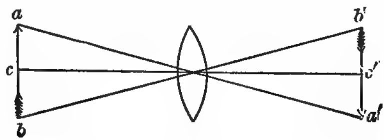

Fig. 106.

ages on secondary axes drawn through those points respectively, and a point called the optical centre of the lens. So long as these secondary axes make but a small angle with the principal axis, definite foci will be formed at the same distances as on the principal axis, and an image $a'b'$ will be formed which will be real and inverted, or virtual and erect, according to the distance of the object from the lens. The formula $\frac{1}{p} + \frac{1}{p'} = (\mu - 1)\left(\frac{1}{r} + \frac{1}{r'}\right) = \frac{1}{f}$ shows that when p increases p' diminishes, and conversely. It shows also that when p is less than f, p' is negative, and the image virtual. It is plain from the figure that the sizes of image and object are in the ratio of their distances from the lens. Diverging lenses, like diverging mirrors, produce only virtual images smaller than the object.

340. Optical Centre.—It was stated in the last section that the secondary axes of a lens pass through a point called the *optical centre*. The position of this point is determined as follows: In Fig. 107, let C, C' be the centres of curvature of the two surfaces of the lens, and let CA and $C'B$ be two parallel radii. The tangents at A and B are also parallel, and light entering at B and emerging at A is light passing through a medium with parallel surfaces (§ 334), and suffers no deviation.

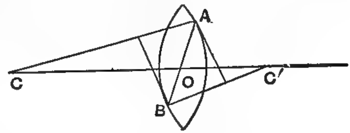

Fig. 107.

If we draw AB, cutting the axis at O, the triangle CAO, $C'BO$ are similar, and $\frac{CA}{C'B} = \frac{CO}{C'O}$. But $\frac{CA}{C'B}$, being the ratio of the radii, is constant for all parts of the surfaces, hence $\frac{CO}{C'O}$ must be constant, or all lines such as AB must cut the axis at one point O. O is the optical centre, and light passing through it is not deviated by the lens.

341. Geometrical Construction of Images.—For the geometrical construction of images formed by curved surfaces, it is convenient to use, in place of the waves themselves, lines perpendicular to the wave front, which represent the paths which the light follows, and are called *rays* of light. These rays, when perpendicular to a plane wave surface, are parallel, and an assemblage of such rays, limited by an aperture in a screen, is called a *beam*. When the rays are perpendicular to a spherical wave surface, they pass through the wave centre, and constitute a *pencil*.

A plane wave surface perpendicular to the axis of a lens is converted by the lens into a spherical wave surface with its centre at the principal focus. The rays perpendicular to the plane wave surface are parallel to the axis, and after emergence must all pass through the principal focus. Conversely, rays emanating from the principal focus emerge from the lens as rays parallel to the axis. Also, rays emanating from any focus must, after emerging from the lens, meet at the conjugate focus. Let L, Fig. 108, be a converging lens, and AB an object. Let O be the optical centre, and F the principal focus. Since all the rays from A must meet, after emerging from the lens, at the conjugate focus, which is the image of A, to find the position of the image it is only necessary to draw two such rays and find their intersection. The ray through the optical centre is not deviated, and the straight line AA' represents both the incident and emergent rays. The ray AL may be consid-

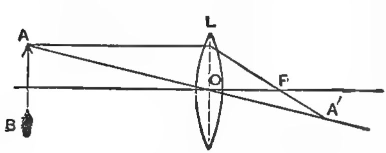

FIG. 108.

ered as one of a group parallel to the axis. All such rays must, after passing through the lens, pass through the principal focus. LA', passing through F, is therefore the emerging ray, and its intersection with AA' locates the image of A. Hence, to construct the image of a point, draw from the point two incident rays, and determine the corresponding emergent rays. The intersection of these will determine the image. The rays most convenient to use are the ray through the optical centre and the ray parallel to the axis or through the principal focus. Fig. 109 gives another example of an image determined by construction.

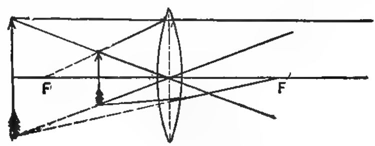

Fig. 109.

342. Thick Lenses.—When a lens is of considerable thickness, the formula derived in § 337 does not give the true position of the conjugate foci. A formula involving the thickness of the lens may be derived without difficulty, but for practical purposes it is usual to refer all measurements to two planes, called the *principal planes* of the lens. The determination of the position of these planes involves a discussion which does not come within the scope of this book.

343. Mirrors and Lenses of Large Aperture.—The equations derived in §§ 336, 337 are only approximations, applying with sufficient exactness to mirrors and lenses of small aperture. But for large apertures, terms containing the higher powers of x cannot be neglected, x will not disappear from the expression of p', and p' will, therefore, not have a definite value. In other words, the reflected or refracted wave is not spherical, and there is no one point l where the light will be concentrated. Surfaces may, however, be constructed which will, in certain particular cases, produce by reflection or refraction perfectly spherical waves. If we desire to find a surface such that light from L (Fig. 110) is concentrated by reflection at l, we remember that the sum $LB + Bl$ must be constant, and that this

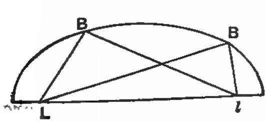

Fig. 110.

is a property of an ellipse with foci at L and l. If the ellipse be constructed and revolved about Ll as an axis, it will generate a surface which will have the required property. If one of the points L be removed to an infinite distance, the corresponding wave becomes a plane perpendicular to Ll, and we must have $LB + BC$ (Fig. 111) constant, a property of the parabola. A parabolic mirror will therefore concentrate at its focus incident light moving in paths parallel to its axis, or will reflect incident light diverging from its focus in plane waves perpendicular to its axis.

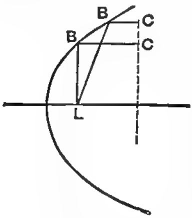

FIG. 111.

Mirrors and lenses having surfaces which are not spherical are seldom made because of mechanical difficulties of construction. It becomes necessary, therefore, to consider how the disadvantages arising from the use of spherical surfaces of large aperture for reflecting or refracting light may be avoided or reduced.

We will consider first the case of a spherical mirror. It was shown above that light from one focus of an ellipsoid is reflected from the ellipsoidal surface in perfectly spherical waves concentric with the other focus. Let Fig. 112 represent a plane section through the axis of an ellipsoid, and Fca a small incident pencil of light proceeding from the focus F. $F'ac$ is a section of the reflected pencil. It is a property of the ellipse that the normals to the curve bisect the angles formed by lines to the two foci. The normal ae bisects the angle FaF'', and hence in the triangle FaF' we have $\frac{Fa}{F'a} = \frac{Fe}{F'e}$.

If d move toward c, $F'a$ increases and Fa diminishes. Hence, from the above proportion, $F'e$ must increase and Fe diminish; or, the successive normals as we approach the minor axis cut the major axis in points successively nearer the centre of the ellipse. The normals produced will therefore meet each other at n beyond the axis. If ac be taken small enough, it may be considered the arc of a circle of which an, cn are radii and n the centre. It is there-

fore a meridian section of an element of a spherical surface of which Fn is an axis.

Sections of wave surfaces reflected from the ellipsoid have their centre at F', and are also sections of wave surfaces reflected from the elementary spherical surface. Evidently the same would be true for any other meridian section passing through FA of the

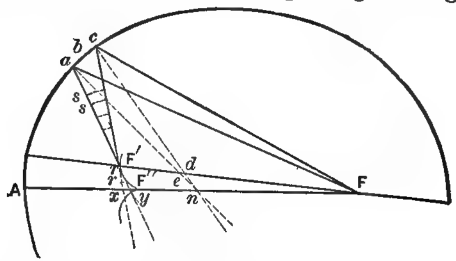

FIG. 112.

sphere of which the elementary surface forms a part, and the form of the wave surfaces may be conceived by supposing the whole figure to revolve about FA as an axis. The arc ac describes a zone of the sphere, s, s, r, r, describe wave surfaces, and F' describes a circumference having its centre on FA. The wave surfaces are portions of the surfaces of curved tubes of which the axis is the arc described by the point F'. The line described by F' is a *focal line,* and all the light from the zone described by ac passes through it, or does so very approximately. If ac be taken nearer to A on the sphere, F' approaches the axis along the curve $F'F''$ and finally coincides with F'', the focus conjugate to F. $F'F''$ is a *caustic curve,* which, when the figure revolves about the axis AF, describes a *caustic surface.* It will be noted that all the light from the zone described by ac passes through the axis AF between the points x and y. The light coming from F and reflected from a small portion of the spherical surface around b, the middle point of ac, is then concentrated first in a line through F' at right angles to the paper, and again into the line xy in the plane of the paper. Nowhere is it concentrated into a point. A line drawn through b and the middle of the focal line through F' is the axis of the re-

flected pencil. It will intersect the axis of the mirror between x and y. If a plane be passed through the point of intersection perpendicular to the axis of the pencil, its intersection with the pencil will be like an elongated figure 8, which may be considered as a focal line at right angles to the axis of the pencil, and in the plane of the paper, and therefore at right angles to the focal line through F'. Between these two focal lines there is a section of least area, nearly circular, which is the nearest approach to an image of F produced by an oblique incidence such as we have been considering.

If refraction instead of reflection had taken place at ac, a result very similar would have been obtained for the refracted pencil. This failure of spherical reflecting or refracting surfaces to bring the light exactly to a focus is called *spherical aberration.* In order to obtain a sharp focus, therefore, if only a single spherical surface be employed, the light must be confined within narrow limits of normal incidence. When reflection or refraction takes place at two or more surfaces in succession, the aberration of one may be made to partially correct the aberration of the other. For instance, when the waves incident upon a double convex lens are plane, the emerging waves are most nearly spherical when the radius of the second surface is six times that of the first. Two or more lenses may be so constructed and combined as to give, for sources of light at a certain distance, almost perfectly spherical emerging waves. Such combinations are called *aplanatic.* The same term is applied to single surfaces so formed as to give by reflection or refraction truly spherical waves.

SIMPLE OPTICAL INSTRUMENTS.

344. The Camera Obscura.—If a converging lens be placed in an opening in the window-shutter of a darkened room, well-defined images of external objects will be formed upon a screen placed at a suitable distance. This constitutes a *camera obscura.* The photographer's camera is a box in one side of which is a lens so adjusted as to form an image of external objects on a plate on the opposite side. The relation deduced in § 339 serves to determine the size of

the image which a given lens will produce, or the focal length of a lens necessary to produce an image of a certain size.

345. The Eye as an Optical Instrument.—The eye, as may be seen from Fig. 113, which represents a section by a horizontal plane, is a camera obscura. *a* is a transparent membrane called the cornea, behind which is a watery fluid called the aqueous humor, filling the space between the cornea and the crystalline lens. Behind this is the vitreous humor, filling the entire posterior cavity of the eye. The aqueous humor, crystalline lens, and vitreous humor constitute a system of lenses, equivalent to a single lens of about two and a half centimetres focus, which produces a real inverted image of external objects upon a screen of nervous tissue called the retina, which lines the inner surface of the posterior half of the eyeball. The retina is an expansion of the optic nerve. The light that forms the image upon it excites the ends of the nerve, and, through the nerve-fibres leading to the brain, produces a mental impression, which, partly by the aid of the other senses, we have learned to interpret as the characteristics of the object the image of which produces the impression. For distinct vision the image must be sharply formed on the retina; but as an object approaches, its image recedes from a lens, and if, in the eye, there were no compensation, we could see distinctly objects only at one distance. The eye, however, adjusts itself to the varying distances of the object by changing the curvature of the front surface of the crystalline lens. There is a limit to this adjustment. For most eyes, an object nearer than fifteen centimetres does not have a distinct image on the retina.

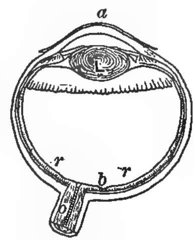

FIG. 113.

We may here consider the means by which we estimate the distance and size of an object. The retina is not all equally sensitive. The depression at *b*, called the *yellow spot*, is much more sensitive than the other portions, and a minute area in the centre of that depression is much more sensitive than the rest of the yellow spot. That part of an image which falls on this small area

is much more distinct than the other parts. How small this most sensitive area is, can be judged by carefully analyzing the effort to see distinctly the minute details of an object. For instance, in looking at the dot of an *i*, a change can be detected in the effort of the muscles that control the eyeball, when the attention is directed from the upper to the lower edge of the dot. The eye can then be directed with great precision to a very small object. The line joining the centre of the crystalline lens with the centre of the sensitive spot may be called the optic axis; and when the attention is directed to any particular point of an object, the eyeballs are turned by a muscular effort, until both the optic axes produced outward meet at the point. For objects at a moderate distance we have learned to associate a particular muscular effort with a particular distance, and our judgment of such distances depends mainly on this association. The angle between the optic axes when they meet at a point is called the *optic angle.* Our estimate of the size of an object is based on our judgment of its distance, together with the angle which the object subtends at the eye, called the *visual angle.* In Fig. 114, when *ab* is an object and *l* the crystalline lens, α is the visual angle. It is plain that the size of the image on the retina is proportional to the visual angle.

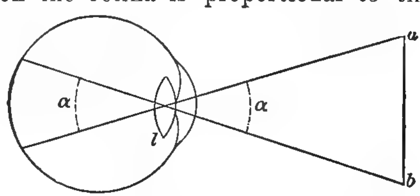

FIG. 114.

It is plain, too, that an object of twice the size, at twice the distance, would subtend the same visual angle and have an image of the same size as *ab*. Nevertheless, if we estimate its distance correctly we shall estimate its size as twice that of *ab*; but if in any way we are deceived as to its distance, and judge it to be less than it really is, we underestimate its size. The visual angle is the *apparent size* of the object.

A less precise estimate of distance can be made with a single

eye, probably from the perception of the effort required to get the object clearly focussed on the retina.

346. Magnifying Power.—To increase the apparent size of an object, and so improve our perception of its details, we must increase the visual angle. This can be done by bringing the object nearer the eye, but it is not always convenient or possible to bring an object near, and even with objects at hand there is a limit to the near approach, due to our inability to see distinctly very near objects. Certain optical instruments serve to increase the visual angle, and so improve our vision. Instruments for examining small objects, and increasing the visual angle beyond that which the object subtends at the nearest point of distinct vision by the unaided eye, are called *microscopes*. Those used for observing a distant object and enlarging the visual angle under which it is seen at that distance are *telescopes*. In both cases the ratio of the visual angles, as the object is seen with the instrument, and without it, is the *magnifying power*.

347. The Magnifying-glass.—Fig. 115 shows how a converging lens may be employed to magnify small objects. The point a of an object just inside the principal focus F of the lens A is the origin of light-waves which, after passing through the lens, are changed to waves having a centre a' (§ 337) which, when the lens

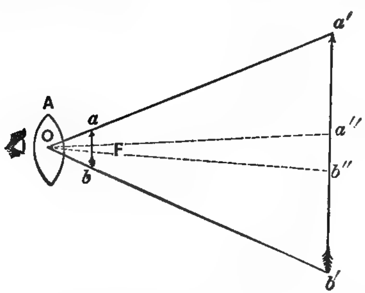

FIG. 115.

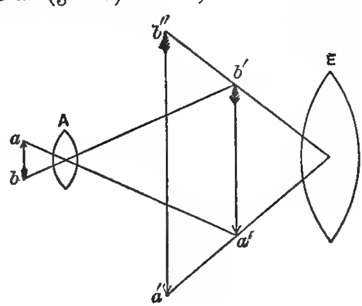

FIG. 116.

is properly adjusted, is at the distance of distinct vision. Waves coming from b enter the eye as though from b'. The object is therefore distinctly seen, but under a visual angle $a'Ob'$, while, to be seen distinctly by the unaided eye, it must be at the distance

Oa'', when the angle subtended is $a''Ob''$. The ratio of these angles is very nearly that of Oa'' to OF. Hence the magnifying power is the ratio of the distance of distinct vision to the focal length of the lens.

348. The Compound Microscope.—A still greater magnifying power may be obtained by first forming a real enlarged image of the object (§ 339) and using the magnifying-glass upon the image, as shown in Fig. 116. The lens A is called the *objective,* and E is called the *eye-lens* or *ocular.* As will be seen in § 359, both A and E often consist of combinations of lenses for the purpose of correcting aberration.

349. Telescopes.—If a lens or mirror be arranged to produce a real image of a distant object, either on a screen or in the air, we may observe the image at the distance of distinct vision, when the visual angle for the object is enlarged in the ratio of the focal length of the lens to the distance of distinct vision. This will be plain from Fig. 117. Suppose the nearest point from which the object can be observed by the naked eye to be the centre of the lens O.

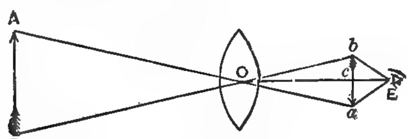

FIG. 117.

The visual angle is then $AOB = aOb$, while the visual angle for the image is aEb. Since these angles are always very small, we have $\frac{aEb}{aOb} = \frac{Oc}{Ec}$ very nearly. But when AB is at a great distance, Oc is the focal length of the lens. By using a magnifying-glass to observe the image, the magnifying power may be still further increased in the ratio of the distance of distinct vision to the focal length of the magnifying-glass. The magnifying power of the combination is therefore the ratio of the focal length of the object-glass to the focal length of the eye-glass. A concave mirror may be substituted for the object-glass for producing the real image.

CHAPTER III.

INTERFERENCE AND DIFFRACTION.

350. Interference of Light from Two Similar Sources.—It has already been shown that the disturbance propagated to any point from a luminous wave is the algebraic sum of the disturbances propagated from the various elements of the wave. The phenomena due to this composition of light-waves are called *interference* phenomena.

Let us consider the case in which two elements only are efficient in producing the disturbance. Let A and B (Fig. 118) represent two elements of the same wave surface separated by the very small distance AB. The disturbance at m, a point on a distant screen mn, parallel with AB, due to these two elements, is the resultant of the disturbances due to each separately. The light is supposed to be homogeneous, and its wave length is represented by λ.

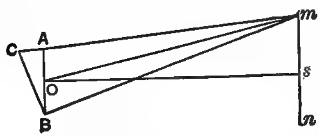

FIG. 118.

When the distance $mB - mA$ equals $\frac{1}{2}\lambda$, or any odd multiple of $\frac{1}{2}\lambda$, there will be no disturbance at m. Take $mC = mB$, and draw BC. mCB is an isosceles triangle; but since AB is very small compared to Om, the angle at C may be taken as a right angle; the triangle ACB, therefore, is similar to Osm, and we have $\frac{AB}{AC} = \frac{Om}{sm} = \frac{Os}{sm}$ very nearly. Represent sm by x, Os by c, AB by b, AC by $n \times \frac{1}{2}\lambda$, where n is any number. Then we have

$$x = \frac{\frac{1}{2}\lambda cn}{b}. \tag{113}$$

If n be any even whole number, the values of x given by this equation represent points on the screen mn at which the waves from A and B meet in the same phase and unite to produce light. If n be any odd whole number, the corresponding values of x represent points where the waves meet in opposite phases, and therefore produce darkness. It appears, therefore, that starting from s, for which $n = 0$, we shall have darkness at distances $\frac{\frac{1}{2}\lambda c}{b}, \frac{\frac{3}{2}\lambda c}{b}, \frac{\frac{5}{2}\lambda c}{b}$, etc., and light at distances $0, \frac{\lambda c}{b}, \frac{2\lambda c}{b}, \frac{3\lambda c}{b}$, etc.

From equation (113) we have $n = \frac{2bx}{c\lambda}$. Since $\frac{1}{2}n\lambda$ is the number of wave lengths that the wave front from B falls behind that from A, $\frac{1}{2}nT$, where T represents the period of one vibration, is the time that must elapse after the wave from A produces a certain displacement before that from B produces a similar displacement. The expression $\frac{2\pi\frac{1}{2}nT}{T} = n\pi$ is, therefore, the difference in epoch of the two wave systems. Substituting $n\pi$ for ϵ in equation (17), we have $S = s + s' = a(2 + 2\cos n\pi)^{\frac{1}{2}} \cos\left(\frac{2\pi t}{T} - \tan^{-1}\frac{\sin n\pi}{1 + \cos n\pi}\right)$. Now the intensity of light for a vibration of any given period is proportional to the mean energy of the vibratory motion, and this can be shown to be proportional to the square of the amplitude. Substituting in the expression for the amplitude the value of n and squaring, we have $A^2 = a^2\left(2 + 2\cos\frac{2bx}{c\lambda}\pi\right)$, in which A^2 is proportional to the intensity of the illumination at distances x from s. When $\frac{2bx}{c\lambda}\pi = 0$, its cosine is 1, and A^2 is a maximum and equal to $4a^2$. As x increases A^2 diminishes, until $\frac{2bx}{c\lambda}\pi = \pi$, in which case $A^2 = 0$. A^2 then increases until it becomes again a maximum, when $\frac{2bx}{c\lambda}\pi = 2\pi$. In short, if AB (Fig. 119) represent the line mn of Fig. 118, the ordinates to a sinuous curve like abc will represent the intensities of the light along that line.

The phenomena described above may be obtained experimentally in several ways. Young admitted sunlight into a darkened room through a small hole in a window-shutter. It fell upon a screen in which were two small holes close together, and, on passing through these, was received upon a second screen. Light and dark bands were observed upon this screen, the distances of which from the central band were in accordance with theory.

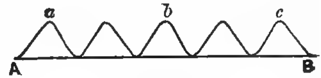

FIG. 119.

Fresnel received the light from a small luminous source upon two mirrors making a very large angle, as in Fig. 120. The light reflected from each mirror proceeded as though from the image of the source produced by that mirror. The reflected light, therefore, consisted of two wave systems, from two precisely similar sources A and B. Light and dark bands were formed in accordance with theory. In order that the experiment may be successfully repeated reflection must take place from the front surface of each mirror only, the angle made by the mirrors must be nearly 180°, and the reflecting surfaces must meet exactly at the vertex of the angle. Two similar sources of light may be obtained also by sending the light through a double prism, as shown in Fig. 121. Light from A proceeds after passing through the prism as from the two virtual images a and a'.

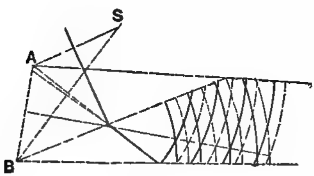

FIG. 120.

FIG. 121.

A divided lens, Fig. 122, serves the same purpose. The light from A is concentrated in two real images a and a', from which proceed two wave systems as in the previous cases. What are really seen in these cases, when the source of light is white, are iris-colored bands instead of bands of light and darkness merely. When the light is monochromatic, the bands are simply alternations of light and darkness, the distances between them being greatest for red light, and least for blue. From equation (113) it appears that,

other things being equal, x varies with λ; hence we must conclude that the greater distance between the bands indicates a greater

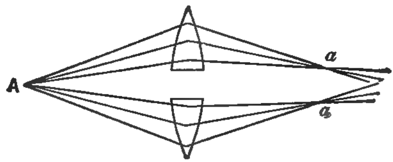

FIG. 122.

wave length; that is, that the wave length of red light is greater than that of blue.

351. Measurement of Wave Lengths.—Data may be obtained from any of the above experiments for the determination of the wave length of light. From equation (113) we have $\lambda = \frac{2bx}{cn}$, where c, b, and x are distances that can be measured. The distance x is the distance from s to a point m, the centre of a light band, and n equals twice the number of dark bands between s and m. Better methods than this of measuring wave lengths will be found described in § 355.

352. Interference from Thin Films.—Thin films of transparent substances, such as the wall of a soap-bubble or a film of oil on water, present interference phenomena when seen in a strong light, due to the interference of waves reflected from the two surfaces of the film. Let AA, BB (Fig. 123) be the surfaces of a transparent film. Light falling on AA is partly reflected and partly transmitted. The reflection at the upper surface takes place with change of sign (§ 132). The light entering the film is partly reflected at the lower surface without change of sign, and returning partly emerges at the upper surface. It is there compounded with the wave at that moment reflected.

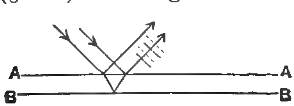

FIG. 123.

Let us suppose the light homogeneous, and the thickness of the film such that the time occupied by the light in going through it and returning is the time of one complete vibration. The returning wave will be in the same phase as the one at that moment entering,

and, therefore, opposite in phase to the wave then reflected. The reflected and emerging waves destroy each other, or would do so if their amplitudes were equal, and the result is that, apparently, no light is reflected. If the light falling on the film be white light, any one of its constituents will be suppressed when the time occupied in going through the film and returning is the period of one vibration, or any whole number of such periods, of that constituent. The remaining constituents produce a tint which is the apparent color of the film.

Similar phenomena are produced by the interference of that portion of the incident light which is transmitted directly through the film, with that portion which is transmitted after undergoing an even number of internal reflections. Since these reflections occur without change of sign, the thickness of the film for which the reflected light is a minimum is that for which the transmitted light is a maximum.

This theory must be slightly modified on account of the internal reflections in the film. The light which enters the film and is reflected does not all pass out in the reflected beam, but part of it is again sent through the film to the other surface, when it is again divided, so that the reflected and transmitted beams both contain light that has been several times reflected. The theory shows that the reflected beam is totally extinguished when the thickness is that indicated by the elementary theory, and that the transmitted beam is never totally extinguished, but merely passes through a minimum intensity. This conclusion is confirmed by observation.

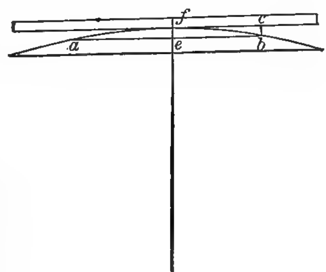

FIG. 124.

Newton was the first to study these phenomena. He placed a plane glass plate upon a convex lens of long radius, and thus formed between the two a film of air, the thickness of which at any point could be determined when the radius of the sphere and the distance from the

point of contact were known. With this arrangement Newton found a black spot at the point of contact, and surrounding this, when white light was used, rings of different colors. When homogeneous light was used, the rings were alternately light and dark. Let ae (Fig. 124) be the radius of the first dark ring, and denote it by d, and let r represent the radius of curvature of the lens. The thickness $bc = ef$, which may be denoted by x, is $x = \frac{d^2}{2r - x}$. Since x is very small in comparison with $2r$, this becomes $x = \frac{d^2}{2r}$. This distance for the first dark ring, when the incident light is normal to the plate, is equal to half the wave length of the light experimented upon. Newton found the thickness for the first dark ring $\frac{2}{178000}$ inches, which corresponds to a wave length of about $\frac{1}{44500}$ inches, or 0.00057 mm. This method affords a means of measuring the wave lengths of light, or, if the wave lengths be known, we may determine the thickness of a film at any point.

353. Effects Produced by Narrow Apertures.—It has been seen (§ 325) that cutting off a portion of a light-wave by means of screens, thus leaving a narrow aperture for the passage of the light, prevents the interference which confines the light to straight lines, and gives rise to a luminous disturbance within the geometrical shadow. This phenomenon is called *diffraction*. Let us consider the aperture perpendicular to the plane of the paper, and an approaching plane wave parallel to the plane of the aperture. Let AB (Fig. 125) represent the aperture, and mn one position of the approaching wave. To determine the effect at any point we must consider the elementary waves proceeding from the various points of the wave front lying between A and B. First consider the point P on the perpendicular to AB at its middle point. AB is so small that the distances from P to each point of AB may be regarded as equal, or

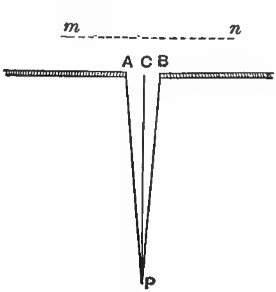

FIG. 125.

the time of passage of the light from each point of AB to P may be made the same, by placing a converging lens of proper focus between AB and P. Then all the elementary waves from points of AB meet at P in the same phase, and the point P is illuminated. Now consider a second point, P', in an oblique direction from C (Fig. 126), and suppose the obliquity such that the time of passage from B to P' is half a vibration period less than the time of passage from C to P', and a whole vibration period less than the time of passage from A to P'. Plainly the elementary waves from B and C will meet at P' in opposite phases, and every wave from a point between B and C will meet at P' a wave in the opposite phase from some point between C and A. The point P' is, therefore, not illuminated. Suppose another point, P'' (Fig. 127), still further from P, such that AB may be divided into three equal parts, each of which is half a wave length nearer P'' than the adjacent part. It is plain that the two parts Bc and ca will annul each other's effects at P'', but that the odd part Aa will furnish light. At a greater obliquity, AB may be divided into four parts, the distances of which from the point, taken in succession, differ by half a wave length. There being an even number of these parts, the sum of their effects at the point will be zero. Now let us suppose the point P to occupy successively all positions to the right or left of the normal. While the line joining P with the middle of the aperture is only slightly oblique, the elementary waves meet at P in nearly the same phase, and the loss of light is small. As P approaches P' (Fig. 126), more and more of the waves meet in opposite phases, the light grows rapidly less, and at P' becomes zero. Going beyond P' the two parts that annul each other's effects no longer occupy the whole space AB, some of the points of the aperture send to P waves

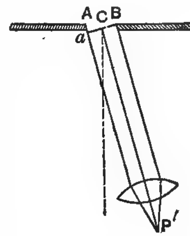

FIG. 126.

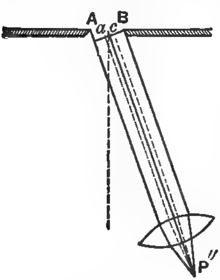

FIG. 127.

that are not neutralized, and the light reappears, giving a second maximum, much less than the first in intensity. Beyond this the light diminishes rapidly in intensity until a point is reached where the paths differing by half a wave length divide AB into four parts, when the light is again zero. Theoretically, maximum and minimum values alternate in this way, to an indefinite distance, but the successive maxima decrease so rapidly that, in reality, only a few bands can be seen.

354. Effect of a Narrow Screen in the Path of the Light.—It can be shown that the effect of a narrow screen is the complement of that of a narrow aperture; that is, where a narrow aperture gives light, a screen produces darkness. Let mn (Fig. 128) be a plane wave and AB a surface on which the light falls. If no obstacle intervene, the surface AB will be equally illuminated. The illumination at any point C is the sum of the effects of all parts of the wave mn. Let the effects due to the part of the wave op be represented by I and that due to all the rest of the wave by I'. Then the illumination at C is $I + I'$, equal to the general illumination on the surface. Let us now suppose mn to be a screen and po a narrow aperture in it. If the illumination at C remain unchanged, it must be that the parts mo and pn of the wave had no effect, and if, for the screen with the narrow aperture, we substitute a narrow screen at op, there will be darkness at C. If, however, a dark band fall at C, when op is an aperture, a screen at op will not cut off the light from C. That is, if C be illuminated when op is an aperture, it will be in darkness when op is a screen; and if it be in darkness when op is an aperture, it will be illuminated when op is a screen.

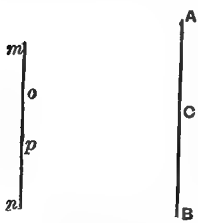

FIG. 128.

355. Diffraction Gratings.—Let AB (Fig. 129) be a screen having several narrow rectangular apertures parallel and equidistant. Such a screen is called a *grating*. Let the approaching waves, moving in the direction of the arrow, be plane and parallel to AB, and let the points a, c, etc., be the centres of the apertures. Draw

the parallel lines ab, cd, etc., at such an angle that the distance from the centre of a to the foot of the perpendicular let fall from the centre of the adjacent opening on ab shall be equal to some definite wave length of light. It is evident that an will contain an exact whole number of wave lengths, co one wave length less, etc. The line mn is, therefore, tangent to the fronts of a series of elementary waves which are in the same phase, and may be considered as a plane wave, which, if it were received on a converging lens, would be concentrated to a focus. If the obliquity of the lines be increased until ae equals 2λ, 3λ, etc., the result will be the same. Let us, however, suppose that ae is not an exact multiple of a wave length, but some fractional part of a wave length, $\frac{99}{100}\lambda$ for example. Let m be the fifty-first opening counting from a; then an will be $\frac{99}{100}\lambda \times 50 = 49.5\lambda$. Hence the wave from the first opening will be in the opposite phase to that from the fifty-first. So the wave from the second opening will be in the opposite phase to that from the fifty-second, etc. If there were one hundred openings in the screen, the second fifty would exactly neutralize the effect of the first fifty in the direction assumed. Light is found, therefore, only in directions given by

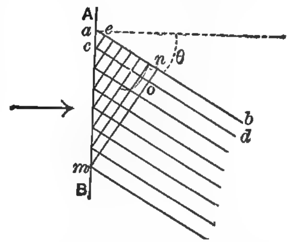

FIG. 129.

$$\sin \theta = \frac{n\lambda}{d}, \tag{114}$$

where n is a whole number, θ the angle between the direction of the light and the normal to the grating, and d the distance from centre to centre of the openings, usually called an *element* of the grating. Gratings are made by ruling lines on glass at the rate of some thousands to the centimetre. The rulings may also be made on the polished surface of speculum metal, and the same effects as described above are produced by reflection from its surface. Since the number of lines on one of these gratings is several thousands, it is seen that the direction of the light is closely confined to the

direction given by the formula, or, in other words, light of only one wave length is found in any one direction. If white light, or any light consisting of waves of various lengths, fall on the grating, the light corresponding to different wave lengths will make different angles with AC, that is, the light is separated into its several constituents, and produces a *pure spectrum*. Since different values of n will give different values of θ for each value of λ, it is plain that there will be several spectra corresponding to the several values of n. When n equals 1 the spectrum is of the *first order*; when n equals 2 the spectrum is of the *second order*, etc. The grating furnishes the most accurate and at the same time the most simple method of determining the wave lengths of light. Knowing the width of an element of the grating, it is only necessary to measure θ for any given kind of light.

Hitherto the spaces from which the elementary waves proceed have been considered infinitely narrow, so that only one system of waves from each space need be considered. In practice, these spaces must have some width, and it may happen that the waves from two parts of the same space may cancel each other. Let the openings, Fig. 130, be equal in width to the opaque spaces, and let the direction am be taken such that ae equals 2λ. Then ae' equals $\frac{1}{2}\lambda$, or the waves from one half of each opening are opposite in phase to those from the other half, and there can be no light in the direction am. In general, if d equal the width of the opening, there will be interference, and light will be destroyed in that direction for which $\sin\theta = \frac{n\lambda}{d}$, if the incident light be normal to the grating. Let f represent the width of the opaque space. Then $d + f = s$, and light occurs in the direction given by $\sin\theta = \frac{n\lambda}{d+f}$, provided that the value of θ given by this equation does not satisfy the first equation also.

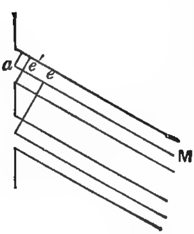

FIG. 130.

If d equal f, we have $\sin\theta = \frac{n\lambda}{d+f} = \frac{n\lambda}{2d}$. When n is even,

$\sin\,\theta$ becomes $\frac{2\lambda}{2d}=\frac{\lambda}{d}$; $\frac{4\lambda}{2d}=\frac{2\lambda}{d}$, etc., and satisfies the equation $\sin\,\theta=\frac{n\lambda}{d}$, which expresses the condition under which light is all destroyed. Hence in this case all the spectra of even orders fail. Moreover, the spectra after the first are not brilliant. When f equals $2d$ the spectrum of the third order fails.

356. Measurement of Wave Lengths.—To realize practically the conditions assumed in the theoretical discussion of the last section, some accessory apparatus is required. It has been assumed that the wave incident upon the grating was plane. Such a wave would proceed from a luminous point or line at an infinite distance. In practice it may be obtained by illuminating a very narrow slit, taking it as the source of light, and placing it in the principal focal plane of a well-corrected converging lens. The plane wave thus obtained passes through the grating, or is reflected from it, and is received on a second lens similar to the first, which gives an image either on a screen or in front of an eyepiece, where it is viewed by the eye. The general construction of the apparatus may be inferred from Fig. 131. It is called the *spectrometer*.

A is a tube carrying at its outer end the slit and at its inner end the lens, called a *collimating* lens. *CD* is a horizontal graduated circle, at the centre of which is a table on which the grating is mounted, and so adjusted that the axis of the circle lies in its plane and parallel to its lines. In using a reflecting grating the collimating and observing telescopes may be fixed at a constant angle 2β with each other, which may be determined once for all in making the adjustments of the instrument. To determine this angle the grating is turned until light thrown through the observing telescope upon the grating is reflected back on itself. The position of the graduated circle is then read. The difference between this

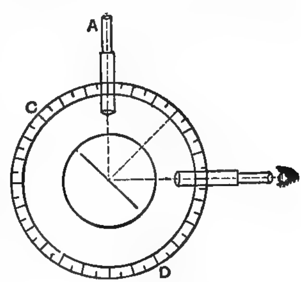

FIG. 131.

reading and the reading when the grating is in such a position that the reflected image of the slit is seen in the telescope is the angle β. If the grating be now turned until the light of which the wave length is required is observed, the angle through which it is turned from its last position is the angle θ. If the width s of an element of the grating be known, these measurements substituted in the equation

$$\lambda = 2s \cos \beta \sin \theta \qquad (115)$$

give the value of λ.

Wave lengths are generally given in terms of a unit called a *tenth metre;* that is, 1 metre $\times 10^{-10}$. The wave lengths of the visible spectrum lie between 7500 and 3900 tenth metres. Langley has found in the lunar radiations wave lengths as long as 170,000 tenth metres, and Rowland has obtained photographs of the solar spectrum in which are lines representing wave lengths of about 3000 tenth metres.

Instead of the arrangement which has been described, Rowland has devised a grating ruled on a concave surface, and is thus enabled to dispense with the collimating lens and the telescope.

CHAPTER IV.

DISPERSION.

357. Dispersion.—When white light falls upon a prism of any refracting medium, it is not only deviated from its course but separated into a number of colored lights, constituting an image called a *spectrum*. These merge imperceptibly from one into another, but there are six markedly different colors: red, orange, yellow, green, blue, and violet. Red is the least and violet the most deviated from the original course of the light. Newton showed by the recomposition of these colors by means of another prism, by a converging lens, and by causing a disk formed of colored sectors to revolve rapidly, that these colors are constituents of white light, and are separated by the prism because of their different refrangibilities. To arrive at a clear understanding of the formation of this spectrum, let us suppose first a small source of homogeneous light, L (Fig. 132). If this light fall on a converging lens from a point at a distance from it a little greater

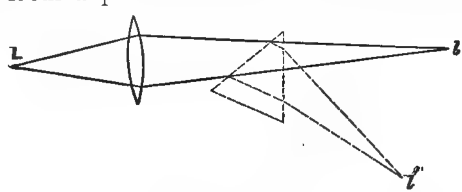

FIG. 132.

than that of the principal focus, a distinct image of the source will be formed at the distant conjugate focus l. If now a prism be placed in the path of the light, it will, if placed so as to give the minimum deviation, merely deviate the light without interfering with the sharpness of the image, which will now be formed at l' instead of at l. If the source L give two or three kinds of

light, the lens may be so constructed as to produce a single sharp image at l of the same color as the source, but when the prism is introduced the lights of different colors will be differently deviated, and two or three distinct images will be found near l'. If there be many such images, some may overlap, and if there be a great number of kinds of light varying progressively in refrangibility, there will be a great number of overlapping images constituting a continuous spectrum.

358. Dispersive Power.—It is found that prisms of different substances giving the same mean deviation of the light deviate the light of different colors differently, and so produce a longer or shorter spectrum. The ratio of the difference between the deviations of the extremities of the spectrum to the mean deviation may be called the *dispersive power* of the substance. Thus if d', d'' represent the extreme deviations, and d the mean deviation, the dispersive power is $\frac{d' - d''}{d}$.

359. Achromatism.—If in Newton's experiment of recomposition of white light by the reversed prism the second prism be of higher dispersive power than the first, and of such an angle as to effect as far as possible the recomposition, the light will not be restored to its original direction, but will still be deviated, and we shall have deviation without dispersion. This is a most important fact in the construction of optical instruments. The dispersion of light by lenses, called *chromatic aberration,* was a serious evil in the early optical instruments, and Newton, who did not think it possible to prevent the dispersion, was led to the construction of reflecting telescopes to remedy the evil. It is plain, however, from what has been said above, that in a combination of two lenses of different kinds of glass, one converging and the other diverging, one may correct the dispersion of the other within certain limits, while the combination still acts as a converging lens forming real images of objects. Fig. 133 shows how this principle is applied to

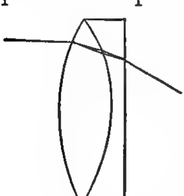

Fig. 133.

the correction of chromatic aberration in the object-glasses of telescopes.

Thus far nothing has been said of the relative separation of the different colors of the spectrum by refraction by different substances. Suppose two prisms of different substances to have such refracting angles that the spectra produced are of the same length. If these two spectra be superposed, the extreme colors may be made to coincide, but the intermediate colors do not coincide at the same time for any two substances of which lenses can be made. Perfect achromatism by means of lenses of two substances is therefore impossible. In practice it is usual to construct an achromatic combination to superpose, not the extreme colors, but those that have most to do with the brilliancy of the image.

The indistinctness due to chromatic aberration, existing even in the compound objective, may be much diminished by a proper disposition of the lenses of the eyepiece. Fig. 134 shows the *negative* or Huygens eyepiece.

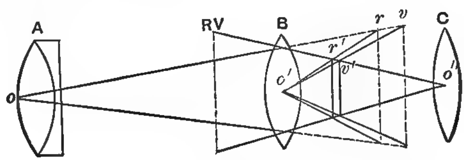

FIG. 134.

Let A be the objective of a telescope or microscope. A point situated on the secondary axis ov would, if the objective were a single lens, have images on that axis, the violet nearest and the red farthest from the lens. If the lens could be perfectly corrected, these images would all concide. By making the lens a little over-corrected, the violet may be made to fall beyond the red. Suppose r and v to be the images. B and C are the two lenses of the Huygens eyepiece. B is called the *field-lens,* and is three times the focal length of C. It is placed between the objective and its focal plane, and therefore prevents the formation of the images rv, but will form images at $r'v'$ on the secondary axes $o'r$, $o'v$. If everything is properly proportioned, $r'v'$ will fall on

the secondary axis $o''R$ of the eye-lens C at such relative distances as to produce *one* virtual image at RV. It will be noted that the image r' is smaller than would have been formed by the objective. The magnifying power of the instrument is therefore less than it would be if the lens C were used alone as the eyepiece. This loss of magnifying power is more than counterbalanced by the increased distinctness.

Fig. 135 shows the *Ramsden* or *positive* eyepiece. The focal length of the lenses in this combination is generally the same, and the distance between them is two-thirds the focal length. The aid it gives in correcting the residual errors of the objective is evident from the figure.

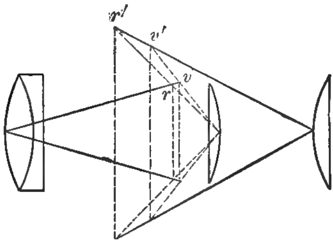

FIG. 135.

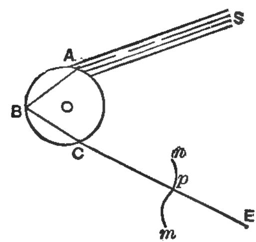

FIG. 136.

360. The Rainbow.—The rainbow is due to refraction and dispersion of sunlight by drops of rain. The complete theory of the rainbow is too abstruse to be given here, but a partial explanation may be given. Let O, Fig. 136, represent a drop of water, and SA the paths of the incident light from the sun. The light enters the drop, suffers refraction on entrance, is reflected from the interior surface near B, and emerges near C, as a wave of double curvature of which mn may be taken as the section. Of this wave the part near p, the point of inflection, gives the maximum effect at a distant point, and if the eye be placed in the prolongation of the line CE perpendicular to the wave surface, light will be perceived, but at a very little distance above or below CE there will be darkness. The direction CE is very nearly that of the minimum deviation produced by the drop with one internal

reflection. It is also the direction in which the angle of emergence equals the angle of incidence. The direction CE corresponds to the minimum deviation for only one kind of light. If this be red light, the yellow will be more deviated, and the blue still more. To see these colors the eye must be higher up, or the drop lower down. If the eye remain stationary, other drops below O will send to it the yellow and blue, and other colors of the spectrum. Since this effect depends only on the angle between the directions SA and CE, it is clear that a similar effect will be received by the eye at E from all drops lying on the cone swept out by the revolution of the line CE and all similar lines drawn to the drops above and below the drop O, about an axis drawn through the sun and the eye, and hence parallel to SA. This cone will trace out the primary rainbow having the red on the outer and the blue on the inner edge. The secondary bow, which is fainter, and appears outside the primary, is produced by two reflections and refractions as shown in Fig. 137.

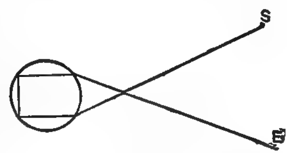

FIG. 137.

361. The Solar Spectrum.—As has been stated (§ 357), solar light when refracted by a prism gives in general a continuous spectrum. Wollaston, in 1802, was the first to observe that when solar light is received upon a prism through a very narrow opening at a considerable distance, dark lines are seen crossing the otherwise continuous spectrum. Later, in 1814–15, Fraunhofer studied these lines, and mapped about 600 of them. That these may be well observed in the prismatic spectrum it is important that the apparatus should be so constructed as to avoid as far as possible spherical and chromatic aberrations. The slit must be very narrow, so that its images may overlap as little as possible. The most important condition for avoiding spherical aberration is that the waves reaching the prism should be plane waves, since all others are distorted by refraction at a plane surface. Fig. 138 shows the disposition of the essential parts of the apparatus known as the *spectroscope*. S is the slit, which may be considered

as the source of light. C is an achromatic lens, called a *collimating* lens, so placed that S is in its principal focus. The waves emerging from it will then be plane. These will be deviated by the prism, and the waves representing the different colors will be

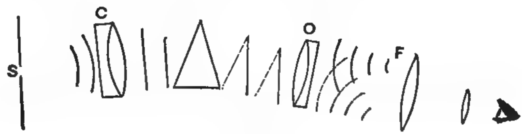

FIG. 138.

separated, so that after passing through the second lens O these different colors will each give a separate image. These images may be received upon a screen, or observed by means of an eye-piece. Sometimes a series of prisms is used to cause a wider separation of the different images.

If the images at F be received on a sensitive photographic plate, it will be found that the image extends far beyond the visible spectrum in the direction of greater refrangibility, and a thermopile or bolometer will show that it also extends a long distance in the opposite direction beyond the visible red. The solar radiations, therefore, do not all have the power of exciting vision. Much the larger part of the solar beam manifests its existence only by other effects. It will be shown that, physically, the various constituents into which white light is separated by the prism differ essentially only in wave length.

CHAPTER V.

ABSORPTION AND EMISSION.

362. Effects of Radiant Energy.—It has been stated that the solar spectrum, whether produced by means of a prism or by a grating, may, under certain conditions, give rise to heat, light, or chemical changes. It was formerly supposed that these were due to three distinct agents emanating from the sun, giving rise to three spectra which were partially superposed. Numerous experiments show, however, that, at any place in the spectrum where light, heat, and chemical effects are produced, nothing which we can do will separate one of these effects from the others. Whatever diminishes the light at any part of the spectrum diminishes the heat and chemical effects also. Physicists are now agreed that all these phenomena are due to vibratory motions transmitted from the sun, which differ in length of wave, and which are separated by a prism, because waves differing in length are transmitted in the substance of the prism with different velocities. The effect produced at any place in the spectrum depends upon the nature of the surface upon which the radiations fall. On the photographic plate they produce chemical change, on the retina the sensation of light, on the thermopile the effect of heat. Only those waves of which the wave lengths lie between 3930 and 7600 tenth metres affect the optic nerve. Chemical changes and the effects of heat are produced by radiations of all wave lengths.

To produce any effect the radiations must be absorbed; that is, the energy of the ethereal vibrations must be imparted to the substance on which they fall, and cease to exist as radiant energy. The

most common effect of such absorption is to generate heat, and there are some surfaces upon which heat will be generated by the absorption of ethereal waves of any length. Langley, by means of the bolometer, has been able to measure the energy throughout the spectrum. He has demonstrated the existence, in the lunar spectrum, of waves as long as 170,000 tenth metres, or more than twenty-two times as long as the longest that can excite human vision.

363. Intensity of Radiations.—The intensity of radiations can only be determined by their effects. If the radiations fall on a body by which they are completely absorbed and converted into heat, the amount of heat developed in unit time may be taken as the measure of the radiant energy. Let us suppose the radiations to emanate from a point equally in all directions, and represent the total energy in a wave by E. Let the point be at the centre of a hollow sphere, of which the radius is r, and represent by I the energy per unit area of the sphere. Then, since the surface of the sphere equals $4\pi r^2$, we have $E = 4\pi r^2 I$,

and
$$I = \frac{E}{4\pi r^2}. \tag{116}$$

That is, the energy which falls upon a given surface is in the inverse ratio of the square of its distance from the source. As we know by experiment that the intensity of light follows the same law, we conclude that the intensity and energy are proportional.

If the surface be not normal to the rays, the radiant energy it receives is less, as will appear from Fig. 139. Let ab be a surface the normal to which makes with the ray the angle θ; then ab will receive the same quantity of radiant energy as $a'b'$, its projection on the plane normal to the ray. But $a'b'$ equals $ab \cos\theta$; and if I represent the energy on unit area of $a'b'$, and I'

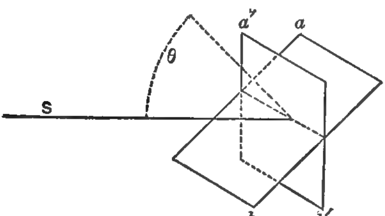

FIG. 139.

the energy on unit area of *ab*, we have $I' = I \cos \theta$; or, the energy of the radiations falling on a given surface is proportional to the cosine of the angle made by the surface and the plane normal to the direction of the rays.

364. Photometry.—The object of *photometry* is to compare the luminous effects of radiations. It is not supposed that the radiations which fall on the retina are totally absorbed by the nerves that impart the sensation of light. The luminous effects, therefore, depend on the susceptibility of these nerves, and can only be compared, at least when different wave lengths are concerned, by means of the eye itself. The photometric comparison of two luminous sources is effected by so placing them that the illuminations produced by them respectively, upon two surfaces conveniently placed for observation, appear to the eye to be equal. If E and E' represent the intensities of the sources, I and I' the intensities of the illuminations produced by them on surfaces at distances r and r', the ratio between these intensities, as was seen in the last section, is

$$\frac{I}{I'} = \frac{\frac{E}{r^2}}{\frac{E'}{r'^2}} = \frac{Er'^2}{E'r^2};$$

and when I and I' are equal, $Er'^2 = E'r^2$, or

$$\frac{E}{E'} = \frac{r^2}{r'^2}. \tag{117}$$

That is, when two luminous sources are so placed as to produce equal illuminations on a surface, their intensities are as the squares of their distances from the illuminated surfaces.

In *Bunsen's photometer* the sources to be compared are placed on the opposite sides of a paper screen, a portion of which has been rendered translucent by oil or paraffine. When this screen is illuminated upon one side only, the translucent portion appears darker on that side, and lighter on the other side, than the opaque portion. When placed between two luminous sources, both sides

of it may, by moving it toward one or the other, be made to appear alike, and the translucent portion almost invisible. The light transmitted through this portion in one direction then equals that transmitted in the opposite direction; that is, the two surfaces are equally illuminated.

365. Transmission and Absorption of Radiations.—It is a familiar fact that colored glass transmits light of certain colors only, and the inference is easy that the other colors are absorbed by the glass. It is only necessary to form a spectrum, and place the colored glass in the path of the light either before or after the separation of the colors, to show which colors are transmitted, and which absorbed.

By the use of the thermopile or bolometer, both of which are sensitive to radiations of all periods of vibration, it is found that some bodies are apparently perfectly transparent to light, and opaque to the obscure radiations. Clear, white glass is opaque to a large portion of the obscure rays of long wave length. Water and solution of alum are still more opaque to these rays, and pure ice transmits almost none of the radiations of which the wave lengths are longer than those of the visible red. Rock salt transmits well both the luminous and the non-luminous radiations.

On the other hand, some substances apparently opaque are transparent to radiations of long wave length. A plate of glass or rock salt rendered opaque to light by smoking it over a lamp is still as transparent as before to the radiations of longer wave length. Selenium is opaque to light, but transparent to the radiations of longer wave length. This fact explains the change of its electrical resistance by light, but not by non-luminous rays. Carbon disulphide, like rock salt, transmits nearly equally the luminous and non-luminous rays; but if iodine be dissolved in it, it will at first cut off the luminous rays of shorter wave length, and as the solution becomes more and more concentrated the absorption extends down the spectrum to the red, and finally all light is extinguished, and the solution to the eye becomes opaque. The radiations of which the wave lengths are longer than those of the red still pass

freely. Black vulcanite seems perfectly opaque, yet it also transmits radiations of long wave length. If the radiations of the electric lamp be concentrated by means of a lens, and a sheet of black vulcanite placed between the lamp and the lens, bodies may be still heated in the focus.

366. Colors of Bodies.—Bodies become visible by the light which comes from them to the eye, and bodies which are not self-luminous must become visible by sending to the eye some portion of the light that falls on them. Of the light which falls on a body, part is reflected from the surface; the remainder which enters the body is, in general, partly absorbed, and the unabsorbed portion either goes on through the body, or is turned back by reflection at a greater or less depth within the body, and mingles with the light reflected from the surface.

In general the surface reflection is small in amount, and the different colors are reflected almost in the proportion in which they exist in the incident light. Much the larger portion of the light by which a body becomes visible is turned back after penetrating a short distance beneath the surface, and contains those colors which the substance does not absorb. This determines the color of the object. In many instances there is a selective reflection from the surface (§ 373). For example, the light reflected from gold-leaf is yellow, while that which it transmits is green.

367. Absorption by Gases.—If a pure spectrum be formed from the white light of the electric lamp, and sodium vapor, obtained by heating a bit of sodium or a bead of common salt in the Bunsen flame, be placed in the path of the beam, two narrow, sharply defined dark lines will be seen to cross the spectrum in the exact position that would be occupied by the yellow lines constituting the spectrum of sodium vapor. Gases in general have an effect similar to that of the vapor of sodium; that is, they absorb from the light which passes through them distinct radiations corresponding to definite wave lengths, which are always the same as those which would be emitted by the gas were it rendered incandescent.

368. Spectrum Analysis.—If the light of a lamp or of any in-

candescent solid, such as the lime of the oxyhydrogen light or the carbons of the electric lamp, be examined with the spectroscope, a continuous spectrum like that produced by sunlight is seen, but the black lines are absent (§ 361). Solids and liquids give in general only continuous spectra. Gases, however, when incandescent give continuous spectra only very rarely. Their spectra are bright lines which are distinct and separate images of the slit. The number and position of these lines differ with each gas employed. Hence, if a mixture of several gases not in chemical combination be heated to incandescence, the spectral lines belonging to each constituent, provided all be present in sufficient quantity, will be found in the resultant spectrum. Such a spectrum will therefore serve to identify the constituents of a mixture of unknown composition. Many chemical compounds are decomposed into their elements, and the elements are rendered gaseous at the temperature necessary for incandescence. In that case the spectrum given is the combined spectra of the elements. A compound gas that does not suffer dissociation at incandescence gives its own spectrum, which is, in general, totally different from the spectra of its elements.

The appearance of a gaseous spectrum depends in some degree on the density of the gas. When the gas is sufficiently compressed, the lines become broader and lose their sharply defined edges, and if the compression be still further increased the lines may widen until they overlap, and form a continuous spectrum. Some of the dark lines of the solar spectrum are found to coincide in position with the bright lines of certain elements. This coincidence is absolute with the most perfect instruments at our command; and not only so, but if the bright lines of the element differ in brilliancy the corresponding dark lines of the solar spectrum differ similarly in darkness.

The close coincidence of some of these lines was noted as early as 1822 by Sir John Herschel, but the absolute coincidence was demonstrated by Kirchhoff, who also pointed out its significance. Placing the flame of a spirit-lamp with a salted wick in the path

of the solar beam which illuminated the slit of his spectroscope, Kirchhoff found the two dark lines corresponding in position to the two bright lines of sodium to become darker, that is, the flame of the lamp had absorbed from the more brilliant solar beam light of the same color as it would itself emit. The explanation of the dark lines of the solar spectrum is obvious. The light from the body of the sun gives a continuous spectrum like that of an incandescent solid or liquid. Somewhere in its course this light passes through an atmosphere of gases which absorbs from the solar beam such light as these gases would emit if they were self-luminous. Some of this absorption occurs in the earth's atmosphere, but most of it is known to occur in the atmosphere of the sun itself. By comparison of these dark lines with the spectra of various incandescent substances upon which we can experiment, the probable constitution of the sun is inferred.

369. Emission of Radiations.—Not only incandescent bodies, but all bodies at whatever temperature they may be, emit radiations. A warm body continues to grow cool until it arrives at the temperature of surrounding bodies, and then if it be moved to a place of lower temperature, it cools still further. To this process we can ascribe no limit, and it is necessary to admit that the body will radiate heat, and so grow cooler, whatever its own temperature, if only it be warmer than surrounding bodies. But it cannot be supposed that a body ceases to radiate heat when it comes to the temperature of surrounding bodies, and begins again when the temperature of these is lowered. It is necessary, therefore, to assume that all bodies, at whatever temperature they may be, are radiating heat, and that, when any one of them arrives at a stationary temperature, it is, if no change take place within it involving the generation or consumption of heat, receiving heat as rapidly as it parts with it. This is called the principle of *movable equilibrium of temperature*, or *Prevost's law of exchanges.* We know that if a number of bodies, none of which are generating or consuming heat otherwise than in change of temperature, be placed in an enclosure, the walls of which are maintained at a constant temperature,

these bodies will in time all come to the temperature of the enclosure. It can be shown that, for this to be true, the ratio of the emissive to the absorbing power must be the same for all bodies, not only for the sum total of all radiations, but for radiations of each wave length. For example, a body which does not absorb radiations of long wave length cannot emit them, otherwise, if placed in an enclosure where it could only receive such radiations, it would become colder than other bodies in the same enclosure. This is only a general statement of the fact which has been already stated for gases, that bodies absorb radiations of exactly the same kind as those which they emit.

Since radiant energy is energy of vibratory motion, it may be supposed to have its origin in the vibrations of the molecules or atoms of the radiating body. It has been shown that the various phenomena of gases are best explained by assuming a constant motion of their molecules. If the atoms of these molecules should have definite periods of vibration, remaining constant for the same gas through wide ranges of pressure and temperature, this would fully explain the peculiarities of the spectra of gases.

In § 150 it was seen that a vibrating body may communicate its vibrations to another body which can vibrate in the same period, and will lose just as much of its own energy of vibration as it imparts to the other body. Moreover, a body which has a definite period of vibration is undisturbed by bodies vibrating in a period different from its own. This explains fully the selective absorption of a gas. For, if a beam of white light pass through a gas, there are, among the vibrations constituting such a beam, some which correspond in period to those of the molecules of the gas, and, unless the energy of vibration of these molecules is already too great, it will be increased at the expense of the vibrations of the same period in the beam of light. Hence, at the parts in the spectrum where light of those vibration periods would fall, the light will be enfeebled, and those parts will appear, by contrast, as dark lines.

In solids and liquids, the molecules are so constrained in their

movements that they do not vibrate in definite periods. Vibrations of all periods may exist; but if in a given case there were a tendency to one period of vibration more than to another, it is evident that the body would transfer to or receive from another, that is, it would emit or absorb, vibrations of that period more than of any other. Furthermore, a good radiator is a body so constituted as to impart to the medium around it the vibratory motion of its own molecules. But the same peculiarity of structure which fits it for communicating its own motion to the medium when its own motion is the greater, fits it also for receiving motion from the medium when its own motion is the less. Theory, therefore, leads us to the conclusion which experiment has established, that at a given temperature emissive and absorbing powers have the same ratio for all bodies.

370. Loss of Heat in Relation to Temperature.—The loss of heat by a body is the more rapid the greater the difference of temperature between it and surrounding bodies. For a small difference of temperature the loss of heat is nearly proportional to this difference. This law is known as *Newton's law of cooling*. For a large difference of temperature the loss of heat increases more rapidly than the difference of temperature, and depends not merely upon this difference, but upon the absolute temperature of the surrounding bodies. An extended series of experiments by Dulong and Petit led to a formula expressing the quantity of heat lost by a body in an enclosure during unit time. It is $Q = m(1.0077)^{\theta}(1.0077^{t} - 1)$, where θ represents the temperature of the enclosure, t the difference of temperature between the enclosure and the radiating body, both measured in Centigrade degrees, and m a constant depending on the substance, and the nature of its surface.

371. Kind of Radiation as Dependent upon Temperature.—When a body is heated we may feel the radiations from its surface long before those radiations render the body visible. If we continue to raise the temperature, after a time the body becomes red-hot; as the temperature rises still further it becomes yellow, and finally attains a white heat. Even this rough observation indicates that the radiations of great wave length are the principal radiations at the lower

temperature, and that to these are added shorter and shorter wave lengths as the temperature rises. Draper showed that the spectrum of a red-hot body exhibits no rays of shorter wave length than the red, but that as the temperature rises the spectrum is extended in the direction of the violet, the additions occurring in the order of the wave lengths. At the same time the colors previously existing increase in brightness, indicating an increase in energy of the vibrations of longer wave length as those of shorter wave length become visible. Experiments by Nichols on the radiations from glowing platinum show that vibrations of shorter wave length are not altogether absent from the radiations of a body of comparatively low temperature, and he was led to believe that all wave lengths are present in the radiations from even the coldest bodies, but are too feeble to be detected.

With gases, as has been seen, the radiations are apparently confined to a few definite wave lengths, but careful observations of the spectra of gases show that the lines are not defined with absolute sharpness, but fade away, although very rapidly, into the dark background. In many cases the existence of radiations may be traced throughout the spectrum, and it is a question whether the spectra of gases are not after all continuous, only showing strongly marked and sharply defined maxima where the lines occur. In general, increase of temperature does not alter the spectra of gases except to increase their intensity, but there are some cases in which additional lines appear as the temperature rises, and a few cases in which the spectrum undergoes a complete change at a certain temperature. This occurs with those compound gases which suffer dissociation at a certain temperature, and at higher temperatures give the spectra of their elements. When it occurs with gases supposed to be elements it suggests the question whether they are not really compounds, the molecules of which at the high temperature are divided, giving new molecules of which the rates of vibration are entirely different from those of the original body.

372. Fluorescence and Phosphorescence.—A few substances,

such as sulphate of quinine, uranium glass, and thallene, have the property, when illuminated by rays of short wave length, even by the invisible rays beyond the violet, of emitting light of longer wave length. Such substances are *fluorescent.* The light emitted by them, and the conditions favorable to their luminosity, have been studied by Stokes. It appears that the light emitted is of the same character, covering a considerable region of the spectrum, no matter what may be the incident light, provided this be such as to produce the effect at all. The light emitted is always of longer wave length than that which causes the luminosity.

There is another class of substances which, after being exposed to light, will glow for some time in the dark. These are *phosphorescent.* They must be carefully distinguished from such bodies as phosphorus and decaying wood, which glow in consequence of chemical action. Some phosphorescent substances, especially the calcium sulphides, glow for several hours after exposure. All fluorescent bodies are also phosphorescent, but the time during which they remain luminous after the exciting light is removed, is so short that it can generally be detected only by special devices.

373. Anomalous Dispersion.—As has been already stated, there is a class of bodies which show a selective absorption at their surfaces. The light reflected from such bodies is complementary to the light which they can transmit. Kundt, following up isolated observations of other physicists, has shown that all such bodies give rise to an *anomalous dispersion;* that is, the order of the colors in the spectrum formed by a prism of one of these substances is not the same as their order in the diffraction spectrum or in the spectrum formed by prisms of substances which do not show selective absorption at their surfaces. Solid fuchsin, when viewed by reflected light, appears green. In solution, when viewed by transmitted light, it appears red. Christiansen allowed light to pass through a prism formed of two glass plates making a small angle with each other, and containing a solution of fuchsin in alcohol. He found that the green was almost totally wanting in

the spectrum, while the order of the other colors was different from that in the normal spectrum. In the spectrum of fuchsin the colors in order, beginning with the one most deviated, were violet, red, orange, and yellow. Other substances give rise to anomalous dispersion in which the order of the colors is different.

In order to account for these phenomena, the ordinary theory of light is extended by the assumption that the ether and molecules of a body materially interact upon one another, so that the vibrations in a light-wave are modified by the vibrations of the molecules of a transparent body through which light is passing. This hypothesis, in the hands of Helmholtz and Ketteler, has been sufficient to account for most of the phenomena of light.

CHAPTER VI.

DOUBLE REFRACTION AND POLARIZATION.

374. Double Refraction in Iceland Spar.—If refraction take place in a medium which is not isotropic, as has been assumed in the previous discussion of refraction, but eolotropic, a new class of phenomena arises. Iceland spar is an eolotropic medium by the use of which the phenomena referred to are strikingly exhibited. Crystals of Iceland spar are rhombohedral in form, and a crystal may be a perfect rhombohedron with six equal plane faces, each of which is a rhombus. Fig. 140 represents such a

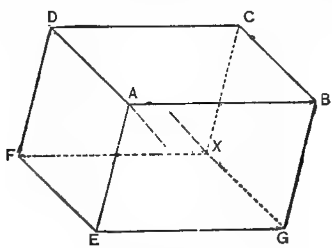

FIG. 140.

crystal. At A and X are two solid angles formed by the obtuse angles of three plane faces. The line through A making equal angles with the three edges $AB, AE, AD,$ or any line parallel to it, is an *optic axis* of the crystal.

Any plane normal to a surface of the crystal and parallel to the optic axis is called a *principal plane.* If such a crystal be laid upon a printed page, the lines of print will, in general, appear double. If a dot be made on a blank paper, and the crystal placed upon it, two images of the dot are seen. If the crystal be revolved about an

axis perpendicular to the paper, one of the images remains stationary, and the other revolves around it. The images lie in a plane perpendicular to the paper, and parallel to the line joining the two obtuse angles of the face by which the light enters or emerges. The entering and emerging light is supposed in this case to be normal to the surfaces of the crystal. If the crystal be turned with its faces oblique to the light, the line joining the images will, in certain cases, not lie parallel to the line joining the obtuse angles of the faces. If the distances of the two images from the observer be carefully noticed it will be seen that the stationary one appears nearer than the other. If the obtuse angles A and X be cut away, and the new surfaces thus formed at right angles to the optic axis be polished, images seen perpendicularly through these faces do not appear double. By cutting the crystal into prisms in various ways its indices of refraction may be measured. It is found that, of the two beams into which light is, in general, divided in the crystal, one obeys the ordinary laws of refraction, and has a refractive index 1.658. It is called the *ordinary ray*. The other has no constant refractive index, does not in general lie in the normal plane containing the incident ray, and refraction may occur when the incidence is normal. It is the *extraordinary ray*. The ratio between the sines of the angles of incidence and refraction varies, for the Fraunhofer line D, from 1.658, the *ordinary index*, to 1.486. This minimum value is called the *extraordinary index*.

375. Explanation of Double Refraction.—In § 334 it was seen that the index of refraction of a substance is the reciprocal of the ratio of the velocity of light in the substance to its velocity in a vacuum. It is plain, then, that the velocity of light for the ordinary ray of the last section is the same for all directions, and that, if light emanate from a point within the crystal, the light, following the ordinary laws of refraction, must proceed in spherical waves about that point as a centre, as in any singly refracting medium. The phenomena presented by the extraordinary light in Iceland spar are fully explained by assuming that the velocities in different directions in the crystal are such as to give a wave front in the

form of a flattened spheroid, of which the polar diameter, parallel to the optic axis, is equal to the diameter of the ordinary spherical

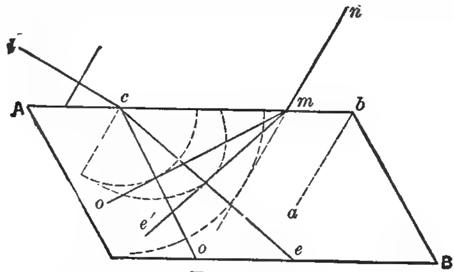

FIG. 141.

wave, and the equatorial diameter is to its polar diameter as 1.658 is to 1.486. From these two wave surfaces the path of the light may easily be determined by construction by methods already explained in § 334, and exemplified in Fig. 141, in which *ic* represents the direction of the incident light, and *co* and *ce* the ordinary and extraordinary rays respectively.

376. Polarization of the Doubly Refracted Light.—If a second crystal be placed in front of the first in any of the experiments described in the last section, there will be seen in general four images instead of two; but if the second crystal be turned, the images change in brightness, and for four positions of the second crystal, when its principal plane is parallel or at right angles to the principal plane of the first, two of the images are invisible, and the other two are at a maximum brightness. If one of the beams of light produced by the first crystal be intercepted by a screen, and the other allowed to pass alone through the second crystal, the phenomena presented are easily followed. If the principal planes of the two crystals coincide, only one image is seen. If the second crystal be now rotated about the beam of light as an axis, a second image at once appears, at first very faint, but increasing in brightness. The original image at the same time diminishes in brightness, and the two are equally bright when the angle between the principal planes is 45°. If the angle be 90° the first image disappears, and the second is at its maximum brilliancy. As the rotation

is continued the first image reappears, while the second grows dim and disappears when the angle between the principal planes is 180°. These changes show that the light which emerges from the first crystal of spar is not ordinary light. Another experiment shows this in a still more striking manner. Let the extraordinary ray be cut off by a screen, and the ordinary ray be received on a plane unsilvered glass at an angle of incidence of 57°. When the plane of incidence coincides with the principal plane of the spar, the light is reflected like ordinary light. If the mirror be now turned about the incident ray as an axis, that is, so turned that, while the angle of incidence remains unchanged, the plane of incidence makes successively all possible angles with the principal plane of the crystal, the reflected light gradually diminishes in brightness, and when the angle between the plane of incidence and the principal plane of the crystal is 90° it fails altogether. If the rotation be continued it gradually returns to its original brightness, which it attains when the angle between the same planes is 180°, and then diminishes until it fails when the angle is 270°. The extraordinary ray presents the same phenomena except that the reflected light is brightest when the angle between the planes is 90° and 270°, and fails when that angle is 0° and 180°. Beams of light after double refraction present different properties on different sides, and are said to be *polarized*. The explanation must, of course, be found in the character of the vibratory motion.

In the polarized beam it is plain that the vibrations must be transverse; for if the light were the result of longitudinal vibrations, or even of vibrations having a longitudinal component, it could not be completely extinguished for certain *azimuths* of the second crystal or of the glass reflector. This conclusion is verified by the experiments of Fresnel and Arago on the interference of polarized light. The difference between ordinary and polarized light is explained if we assume that, in both, the vibrations of the ether particles take place at right angles to the line of propagation of the wave, and that in ordinary light they occur irregularly in all azimuths about that line, and may be performed in ellipses or

circles as well as in straight lines, while in polarized light they occur in one plane. In the ordinary ray in Iceland spar the vibrations are in a plane at right angles to the optic axis. In the extraordinary ray they are in the plane containing the optic axis and the ray. If we assume that the rigidity of the ether is different in different directions in the crystal, that at right angles to the optic axis it is a minimum and along the optic axis a maximum, and varies between these two directions according to a simple law, all the phenomena of double refraction and polarization in the crystal are accounted for. If a crystal be cut so as to present faces parallel to the optic axis, and if light enter along a normal to one of these faces, the vibrations, which previous to entering the crystal were in all azimuths, are resolved in it in two directions—that of greatest and that of least elasticity, or parallel to and at right angles to the optic axis. The wave made up of vibrations parallel to the optic axis is propagated with the greater velocity. In this case the two wave fronts continue in parallel planes, and upon emergence constitute apparently one beam of light. If the incidence be oblique and in a plane at right angles to the principal plane, the two component vibrations are still parallel to and at right angles to the optic axis, but a refraction occurs which is greater for the ray of which the vibrations are in the direction of least elasticity. If the incidence be oblique and in the principal plane, it is evident that there may be a component vibration at right angles to the optic axis, but the other component, since it must be at right angles to the ray, cannot be parallel to the optic axis, and therefore cannot be in the direction of greatest elasticity in the crystal. The second component is, however, in the direction of greatest elasticity in the plane of vibration, which direction is at right angles to the first component. In general, if a ray of light pass in any direction within the crystal, the line drawn at right angles to that direction and to the optic axis, that is, at right angles to the plane determined by the ray and the optic axis, is in the direction of least elasticity. One of the component vibrations is in that direction. A line drawn at right angles to the ray and in the plane formed by

it and the optic axis is in the direction of the greatest elasticity to which any vibration giving rise to that ray of light can correspond. In that direction is the second component vibration. The two component vibrations are therefore always at right angles. One of the components is always at right angles to the optic axis, and hence in the direction of least elasticity. The light resulting from this component always travels with the same velocity whatever its direction, and hence suffers refraction on entering the crystal or emerging from it, according to the ordinary law for single refraction. The other component, being in the plane containing the ray and the optic axis and at right angles to the ray, may make all angles with the optic axis from 0°, when it is in the direction of maximum elasticity and is propagated with the greatest velocity, to 90°, when it is in a direction in which the elasticity is the same as that for the other component, and the entire beam is propagated as ordinary light. Light for which vibrations occur in all azimuths will, on entering the crystal, give rise to equal components, but light already polarized will give rise to components the intensities of which are determined by the law for the resolution of motions. When its own direction of vibration coincides with that of either of the components, the other component will be zero, and only when its vibrations make an angle of 45° with the components can these components be equal. The varying intensities of the two beams into which a polarized beam is divided by a second crystal are thus explained.

377. Polarization by Reflection.—Light reflected from a transparent medium is found in general to be partially polarized, and for a certain angle of incidence the polarization is nearly perfect. This angle is that for which the reflected and refracted rays are at right angles. In Fig. 142 let xy represent the surface of a transparent medium, ab the incident, bc the reflected, and bd the refracted ray. If the angle

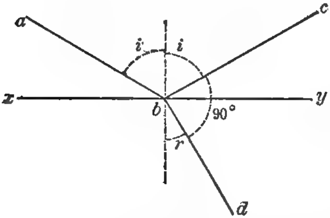

FIG. 142.

$cbd = 90°$, we have $r + i = 90°$ also; and since $\mu = \frac{\sin i}{\sin r}$, we have $\mu = \frac{\sin i}{\cos i} = \tan i$. Hence the angle of complete polarization is given by the equation $\tan i = \mu$. The fact embodied in this equation was discovered by Brewster, and is known as *Brewster's law.* The angle of complete polarization is called the *polarizing angle.* The plane of incidence is the *plane of polarization.* The vibrations of polarized light are assumed to be at right angles to the plane of polarization. In the transmitted ray is an equal amount of polarized light, the vibrations of which are in the plane of incidence.

If a beam of ordinary light traverse a transparent medium, in which are suspended minute solid particles, the light which is reflected from them is found to be partially polarized. The maximum polarization is found in the light reflected at right angles to the beam. The plane of polarization of the polarized beam is the plane of the original beam and the beam which reaches the eye of the observer.

378. Interference of Polarized Light.—The assumption that has been made in the foregoing descriptions, that the vibrations in the polarized beam are transverse to the direction of propagation, is fully justified, not only by the satisfactory way in which it explains the various modes of production of polarized light and the phenomena connected with it, but also by direct experiment. Fresnel and Arago examined the interference of polarized beams and arrived at the following conclusions: Two rays of light polarized at right angles with each other do not appear to affect each other at all in the same circumstances in which two rays of ordinary light destroy each other by interference. Two rays of light polarized in the same plane act on one another like ordinary light, so that in the two cases the phenomena of interference are absolutely the same. Two rays originally polarized at right angles to each other can afterwards be so modified that they are both polarized in the same plane without acquiring the power of interfering with each other. Two rays polarized at right angles and afterwards brought to the same plane of polarization interfere like ordinary light if they come from the same polarized beam.

From the first and second of these statements it is plain that the vibration in the polarized beam must be transverse to the direction of propagation, for if it were otherwise, there would be some interference of the two rays, even when they are polarized at right angles to each other.

We may here consider the nature of *common light*. The peculiarity of common light is that it furnishes two images of equal intensity when it passes through a doubly refracting crystal, and that it cannot produce colored fringes when passed through a crystal plate and examined with an analyzer (§ 379). These peculiarities can be explained by supposing that the direction of vibration in the wave frequently changes. On the other hand, the interference of common light proves that this change of direction does not occur in every wave. In the experiments of Michelson and Morley interference was obtained between two beams of light of which the difference in path was 200,000 wave lengths. Such interference could not have occurred if the direction of the vibration had changed during the time taken by light to traverse that distance. We are accordingly compelled to assume that the vibrations of common light are polarized in one plane for a very short time, which is, however, sufficiently long for the light to execute a large number of vibrations in it, and that at certain intervals the plane of polarization changes its direction.

379. Polariscopes.—In experimenting with polarized light we need a *polarizer* to produce the polarized beam, and an *analyzer* to show the effects of the polarization. A piece of plane glass, reflecting light at the polarizing angle, is a simple polarizer. Doubly refracting crystals, if means be employed to suppress one of the

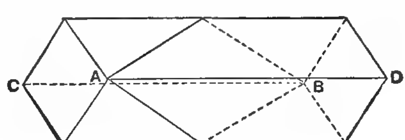

FIG. 143.

beams into which the light is divided, are excellent polarizers. Tourmaline is a doubly refracting crystal which has the property

of being more transparent to the extraordinary than to the ordinary ray. By grinding plates of tourmaline to the proper thickness, the ordinary ray is completely absorbed, while the extraordinary ray is transmitted. The best method of obtaining a polarized beam is by the use of a crystal of Iceland spar in which, by an ingenious device, the ordinary ray is suppressed and the extraordinary transmitted. Fig. 143 shows how this is accomplished. *AB* is a crystal of considerable length. It is divided along the plane *AB*, making an angle of 22° with the edge *AD* and perpendicular to a principal plane of the face *AC*. The faces of the cut are polished and the two halves cemented together again by Canada balsam in the same position as at first. In Fig. 144, which is a section through *ACBD* of Fig. 143, *ab* represents the direction of the light which is incident upon the face *AC*. It is separated into two rays, *o* and *e*. Since the refractive index of the balsam is intermediate between the ordinary and extraordinary indices of the spar, and since the

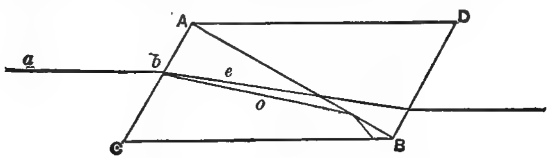

FIG. 144.

angle *DAB* is so chosen that the ray *o* strikes the balsam at an angle of incidence greater than the critical angle, the ray *o* is totally reflected. The ray *e*, on the other hand, having a refractive index in the spar less than in the balsam, is not reflected, but continues through the crystal. A crystal of Iceland spar so treated is called a *Nicol's prism*, or often simply *a Nicol.*

A pair of Nicol's prisms, mounted with their axes coinciding, serves as a *polariscope*. The first Nicol transmits a single beam of polarized light the vibrations of which are in the principal plane. When the principal plane of the second Nicol coincides with that of the first this light is wholly transmitted through it. If the second Nicol or analyzer be turned about its axis, whenever its principal plane makes an angle with the direction of the vibrations, these are resolved into two components, one in and the other at

right angles to the principal plane. The latter is reflected to one side and absorbed, and the former is transmitted. As the angle between the two principal planes increases, the transmitted component diminishes in intensity, until when this angle becomes 90° it disappears entirely. In this position the polarizer and analyzer are said to be *crossed*.

380. Effects of Plates of Doubly Refracting Crystals on Polarized Light.—If a plate cut from a doubly refracting crystal so that its faces are parallel to the optic axis, or at least not at right angles to it, be placed between the crossed polarizer and analyzer, and the principal plane of the plate coincide with, or be at right angles to, the plane of vibration, no effect is perceived. But if the plate be rotated so that its principal plane makes an angle with the plane of vibration, the motion may be considered as resolved into two components, one in, and the other at right angles to, the principal plane of the plate, and these two components on reaching the analyzer are again resolved each into two others, one in, and the other at right angles to, the principal plane of the analyzer. The vibrations in the principal plane of the analyzer are transmitted through it, and hence, in general, the introduction of the plate restores the light which the crossed polarizer and analyzer had extinguished. It is easy to see that the restored light will be most intense when the principal plane of the plate makes an angle of 45° with the plane of vibration of the polarized ray.

It is not to be understood that in the plate there are two separate beams of light, in one of which one set of particles is vibrating in one plane, and in the other another set in another plane. What really takes place is that each particle in the path of the light describes a path which is the resultant of the two components spoken of above. Let ab (Fig. 145) be a plate of Iceland spar, and cd the direction of its optic axis. Suppose the path of the light perpendicular to the plane of the paper, and ef to represent the direction of the disturbance produced by the entrance of a plane polarized wave. A motion in the direction of ef is com-

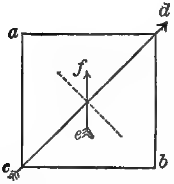

FIG. 145.

pounded of two motions, one along the axis, and the other perpendicular to it. In the propagation of this motion to the next particle, the motion in the direction of the optic axis will begin a little sooner than that at right angles because of the greater elasticity in the former direction, and this difference becomes greater as the light is propagated into the plate. This is equivalent to a change in the relative phases of two vibrations at right angles, and this causes the path of a vibrating particle to change from the straight line to an ellipse. The result is, therefore, that, when the initial disturbance has any direction except in or at right angles to the principal plane of the plate, the motion of the vibrating particles within the plate becomes elliptical, the ellipses changing form as the distance from the front surface of the plate increases. It is entirely admissible, however, in the discussion of the problem to substitute for the actual motion its two components, as was done above.

It remains to consider what is the effect of the retardation or change of phase of one of the components with respect to the other. It will be remembered that in the analyzer each ray from the plate is again resolved into two components, and that two of these components are in the principal plane of the analyzer and are transmitted. These two components will evidently differ in phase just as did the two motions from which they were derived, and since they are in the same plane their resultant is represented by their algebraic sum. If they differ in phase by half a period their algebraic sum will be zero, and no light will be transmitted by the analyzer. This will occur for a certain thickness of the interposed plate. If the light experimented upon be white, it may occur for some wave lengths and not for others. Hence, some of the constituents of white light may fail in the beam transmitted by the analyzer, and the image of the plate will then appear colored. A study of the resolution of the vibrations for this case shows that, of the two beams formed in the analyzer, one contains just that portion of the

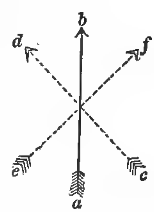

FIG. 146.

light that the other lacks; hence if the analyzer be turned through 90°, the image will change to the complementary color. In Fig. 146, let *ab* represent the plane of the vibrations in the polarized ray, and let *cd* and *ef* represent the two planes of vibration of the rays in the interposed plate. At the instant of entering the plate the primary vibration and its two components will have the relation shown in the figure. The two components are then in the same phase. As the movement penetrates the plate, one component falls behind the other, and the relation of their phases changes, until, with a retardation of one wave length, the phases are again as in the figure. Suppose the thickness of the plate such that this retardation occurs for some constituent of white light. After leaving the plate the relative phases of the components remain unchanged, and the constituent in question enters the analyzer as two vibrations at right angles and in the same phase. In Fig. 147 let *oe* and *od* represent the two components, and *xx* and *yy* the two planes of vibration in the analyzer. *oe* will give the components *om* and *on*, and *od* the components *om′* and *on′*. Since the components *om* and *om′* annul each other, the color to which they correspond is wanting in the light resulting from vibrations in the plane *xx*, while since the components *on* and *on′* are added, this color is found in full intensity among the vibrations in the plane *yy*. For light of other wave lengths, the relative retardation is different, but for each vibration period, the component in the direction *xx* combined with that in the direction *yy* represents the total light for that period in the beam entering the analyzer; that is, the total effect of vibrations in the direction *xx* combined with that of vibrations in the direction *yy* must produce white light, and one effect must, therefore, be the complement of the other.

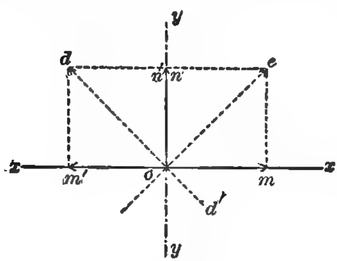

FIG. 147.

Let us suppose the plate thick enough to cause a retardation

equal to a certain number of wave lengths, which we will assume to be ten, of the shortest waves of the visible spectrum. Since the longest waves of the visible spectrum are about twice the length of the shortest, they will suffer a retardation of five wave lengths. Other waves will suffer a retardation of nine, eight, seven, and six wave lengths. But, as was seen above, a retardation of one or more whole wave lengths of any kind of light causes extinction of that kind of light in the beam transmitted by the crossed analyzer. In the case considered the transmitted beam will lose six kinds of light distributed at about equal distances along the spectrum. The light remaining will consist of the different colors in about the same proportions as they exist in white light, and the beam will therefore be white but diminished in intensity. Hence, when a thick plate is interposed between the crossed polarizer and analyzer the restored light is white.

381. Elliptic and Circular Polarization.—In the last section, in discussing the effects of a thin plate, we considered the two components of the vibratory motion propagated from it. It was stated that the real motion of the vibrating particles was in general elliptical. Let us consider more fully the real motion. Let us suppose that the light is light of one wave length only, and that, as before, the principal plane of the plate makes an angle of 45° with the plane of vibration of the incident light. In Fig. 148 let *yy* represent the original plane of vibration, and *ab* and *cd* the planes of maximum and minimum elasticity in the plate. As already explained, the first disturbance as the light enters the plate is in the direction *yy*; but as the disturbance is propagated into the plate, each disturbed particle receives an impulse first of all in the direction *cd* of greatest elasticity, then in other directions between *cd* and *ab*, and finally in the direction *ab*. From this results an elliptical orbit with the major axis in the direction *yy*. To determine this orbit exactly it is

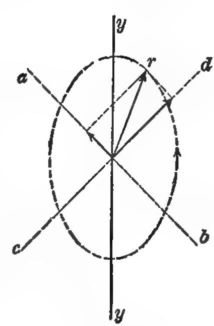

FIG. 148.

only necessary to take account of the time that elapses between the impulse in the direction *cd* and the corresponding impulse in the direction *ab*. It is sufficient to consider any particle as actuated by two vibratory motions in the directions *cd* and *ab* at right angles, and differing in phase. In Fig. 149, one side of the rectangle represents the greatest displacement in the direction *cd*, and the other side the displacement occurring at the same instant in the direction *ba*. The point *r* will represent the actual position of the vibrating particle. Constructing now the successive displacements of the particles in the directions *cd* and *ba* and combining these, we have the elliptical path as shown. As the light penetrates farther and farther into the plate the relative phases of the two vibrations change continually, and the ellipse passes through all its forms from the straight line *yy* to the straight line *xx* at right angles to it and back to the straight line *yy*. The direction of the path of the particle in the surface of the plate as the light emerges will be the direction of the path of all the particles in the polarized beam beyond the plate. If the component vibrations be in the same phase, that is, if they reach their elongations in the directions *ba* and *cd* (Fig. 149) at the same instant, the resultant vibration is in the line *yy* and the light is plane polarized exactly as it left the polarizer. This will occur when the retardation of light in the plane of *ba* with respect to that in the plane of *cd* is one, two, or more whole wave lengths. When the retardation is one half, three halves, or any odd number of half wave lengths, the phases of the two vibrations are as shown in Fig. 150, and the resultant is a plane polarized beam the vibrations of which are at right angles to those of the beam from the polarizer. A case of special interest is shown in Fig. 151, in which the difference of phase is one fourth a period, and the resultant vibration is a circle. A difference of three fourths will give a circle also, but with the rotation in the

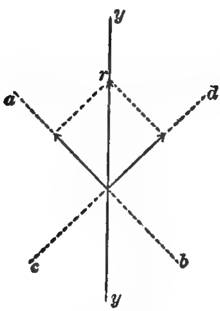

FIG. 149.

opposite direction. A plate of such thickness as to produce a retardation of one quarter of a wave length will give a circular vibration, and the beam issuing from the plate is then *circularly polarized.* Its peculiarity is that the two beams into which it is divided by a doubly refracting crystal are always of the same intensity, and no form of analyzer will distinguish it from ordinary

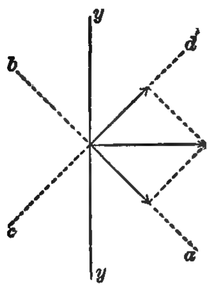

FIG. 150.

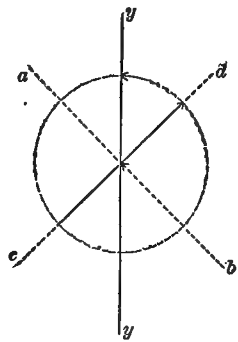

FIG. 151.

light. *Quarter wave plates* are often made by splitting sheets of mica until the required thickness is obtained.

382. Circular Polarization by Reflection.—It has been stated that light reflected from a transparent medium at a certain angle is polarized, and that an equal amount of polarized light exists in the refracted beam. Light totally reflected in the interior of a medium is also polarized, and here, there being no refracted beam, the two components exist in the reflected light, but so related in phase that the light is elliptically polarized. Fresnel has devised an apparatus known as *Fresnel's rhomb,* by means of which circularly polarized light is obtained by two internal reflections of a beam of light previously polarized in a plane at an angle of 45° with the plane of incidence.

383. Effect of Plates Cut Perpendicularly to the Axis from a Uniaxial Crystal.—A crystal, such as Iceland spar, which has but one optic axis, is called a *uniaxial crystal.* Polarized light passing perpendicularly through a plate cut from such a crystal per-

pendicularly to its optic axis suffers no change. If, however, the plate between the crossed polarizer and analyzer be inclined to the direction of the beam, light passes through the analyzer. It is generally colored, the color changing with the obliquity of the plate. If a system of lenses be used to convert the polarized beam into a conical pencil and the plate be placed in this perpendicular to its axis, the central ray of the pencil will be unchanged, but the oblique rays will be resolved except in and at right angles to the plane of vibration, and there will appear beyond the analyzer a system of colored rings surrounding a dark centre, and intersected by a black cross. If the analyzer be turned through 90°, a figure complementary to the first in all its shades and tints is obtained: the black cross and centre become white, and the rings change to complementary colors.

384. Biaxial Crystals.—Most crystals have two optic axes or lines of no double refraction, instead of one. They are *biaxial crystals*. Their optic axes may be inclined to each other at any angle from 0° to 180°. The wave surfaces within these crystals are no longer the sphere and the ellipsoid, but surfaces of the fourth order with two nappes tangent to each other at four points where they are pierced by the optic axes. Neither of the two rays in such a crystal follows the law of ordinary refraction. The outer wave surface around one of the points of tangency has a depression something like that of an apple around the stem. By reference to the method already employed for constructing a wave front, it will be seen that there may be such a position for the incident wave that, when the elementary wave surfaces are constructed, the resultant wave will be a tangent to them in the circle around one of these depressions where it is pierced by the optic axis. Now since the direction of a ray of light is from the centre of an elementary wave surface to the point of tangency of that surface and the resultant wave, we shall have in this case an infinite number of rays forming a cone, of which the base is the circle of tangency. In other words, one ray entering the plate in a proper direction may be resolved into an infinite number of rays

forming a cone, which will become a hollow cylinder of light on emerging from the crystal. This phenomenon is called *conical refraction*. It was predicted by Hamilton from a mathematical analysis of the wave propagation in such crystals.

If a plate be cut from a biaxial crystal perpendicular to the line bisecting the angle formed by the optic axes, and placed between the polarizer and analyzer in a conical pencil of light, there will be seen a series of colored curves called lemniscates, resembling somewhat a figure 8. The existence of this phenomenon was also predicted and the forms of the curves investigated by mathematical analysis before they were seen.

385. Double Refraction by Isotropic Substances when Strained.—A piece of glass between the crossed polarizer and analyzer, if subjected to forces tending to distort it, will restore the light beyond the analyzer and in some cases produce chromatic effects. Unequal heating produces this result, and a long tube made to vibrate longitudinally shows it when the light crosses it near the node. Pieces cut from plates of unannealed glass exhibit double refraction when examined by polarized light. Indeed, the absence of double refraction is a test of perfect annealing.

386. Effects of Plates of Quartz.—A quartz crystal is uniaxial, and gives an ordinary and an extraordinary ray, but is unlike Iceland spar in that the extraordinary wave front in it is a prolate spheroid and lies within the spherical ordinary wave. The effects due to plates of quartz in polarized light differ very greatly from those due to Iceland spar or selenite. If a plate of quartz cut perpendicularly to the axis be placed in a beam of parallel, homogeneous, plane polarized light at right angles to its path, the light is, in general, restored beyond the analyzer, and is unchanged by the rotation of the quartz through any azimuth. If the analyzer be rotated through a certain angle, depending on the thickness of the quartz plate, the light is extinguished. It is evident that the plane of polarization has simply been rotated through a certain angle. Light of a different wave length would have been rotated through a different angle. A beam of white polarized light, therefore, has the

planes of polarization of its constituents rotated through different angles, and the effect of rotating the analyzer is to quench one after another of the colors as the plane of polarization for each is reached. The result is a colored beam which changes its tint continuously as the analyzer rotates.

The best explanation of these phenomena was given by Fresnel. It is found that neither of the two beams from a quartz crystal is plane polarized. The polarization is in general elliptical, but becomes circular for waves perpendicular to the axis of the crystal, the motion in one ray being right-handed and in the other left-handed. Each particle of ether in the path of the light within the crystal is actuated at the same time by two circular motions in opposite directions. Its real motion is in the diameter which bisects the chord joining any two simultaneous hypothetical positions of the particle in the two circles. In Fig. 152 let P and Q represent these two simultaneous positions. It is plain that the two components in the direction AB have the same value and are added, while those at right angles to AB are equal and opposite and annul each other. So long as the two components retain the same relation as that assumed, the real motion of the particle is in the line AB. But in the quartz plate one of the motions is propagated more rapidly than the other, and another particle farther on in the path of the light may reach the point P in one of its circular vibrations at the same time that it reaches Q' in the other. This will give CD as its real path, and the plane of its vibration has been rotated through the angle BOD. When the light finally emerges from the plate its plane of vibration will have been rotated through an angle which is proportional to the thickness of the plate and depends upon the wave length of the light employed. A plate of quartz one millimetre in thickness rotates the plane of polarization of red light corresponding to Fraunhofer's line B, 15° 18′, of blue light corresponding to the line G, 42° 12′. Some specimens of quartz rotate

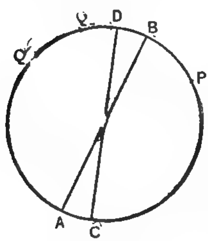

FIG. 152.

the plane of polarization in one direction, and some in the opposite. Rotation which is related to the direction of the light as the directions of rotation and propulsion in a right-handed screw is said to be *right-handed*, and that in the opposite direction is *left-handed*.

Reusch has reproduced all the effects of quartz plates by superposing thin films of mica, each film being turned so that its principal plane makes an angle of 45° or 60°, always in the same direction, with that of the film below. If a plane polarized wave enter such a combination, an analysis of the resolution of the vibration as it passes from film to film will show that the result is equivalent to that of two contrary circular vibrations, one of which is propagated less rapidly than the other. This helps to establish Fresnel's theory of the rotational effects of quartz.

387. Rotation of the Plane of Polarization by Liquids.—Many liquids rotate the plane of polarization, but to a less amount than quartz. A solution of sugar produces a rotation varying with the strength of the solution, and instruments called *saccharimeters* are made for determining the strength of sugar solutions from their effect in rotating the plane of polarization. In these instruments the effect is often measured by interposing a wedge-shaped piece of quartz, and moving it until a thickness is found which exactly compensates the rotation produced by the solution.

388. Electromagnetic Rotation.—Faraday discovered that when polarized light passes through certain substances in a magnetic field the plane of polarization is rotated through a certain angle. The experiment succeeds best with a very dense glass consisting of borate of lead, so placed that the light may traverse it along the lines of magnetic force, in the field produced by a powerful electromagnet. The amount of rotation is proportional to the difference of magnetic potential between the two ends of the glass. The direction of rotation, as was shown by Verdet, is generally right-handed in diamagnetic media, and left-handed in paramagnetic media. It also depends upon the direction of the lines of force, and is therefore reversed by reversing the current in the electromagnet. It follows, also, that if the light, after traversing the glass with the lines of

force, be reflected back through the glass against the lines of force, the rotation will be doubled. It is important to note that this is the reverse of the effect produced by quartz, solutions of sugar, etc., which rotate the plane of polarization in consequence of their own molecular state. When light of which the plane of polarization has been rotated by passage through such substances is reflected back upon itself, the rotation produced during the first passage is exactly reversed during the return, and the returning light is found to be polarized in the same plane as at first.

In the magnetic field the effect is as though the medium which conveys the light were rotating around an axis parallel to the lines of force, and carrying with it the plane of vibration. Evidently the plane of vibration would be turned through a certain angle during the passage of the light through the body, and would be turned still further in the same direction if the light were to return.

When we remember that iron becomes magnetic by the effect of currents of electricity flowing in conductors around it, and that Ampère conceived that a permanent magnet consists of molecules surrounded by electric currents, all in the same direction, it is easy to imagine that the magnetic field is a region where the ether is actuated by vortical motions, all in the same direction, and in planes at right angles to the lines of magnetic force. Such a motion would account for the rotational effects of the magnetic field upon polarized light.

Not only glass but most liquids and gases exhibit rotational effects when placed in a powerful magnetic field; and Kerr has shown that when light is reflected from the polished pole of an electromagnet, its primitive plane of polarization is rotated when the current is passed, in one direction for a north pole, and in the opposite direction for a south pole.

389. Maxwell's Electromagnetic Theory of Light.—In Maxwell's treatment of electricity and magnetism he assumed that electrical and magnetic actions take place through a universal medium. In order to determine whether this medium may not be identical with the luminiferous ether, he investigated its properties when a

periodic electromagnetic disturbance is supposed to be set up in it, such as would result from a rapid reversal of electromotive force at a point, and compared them with the observed properties of the ether, on the assumption that light is an electromagnetic disturbance. He showed that such a disturbance would be propagated through the medium in a way similar to that in which vibrations are transmitted in an elastic solid. He showed further that if light were such a disturbance, its velocity in the ether should be equal to v, the ratio of the electrostatic to the electromagnetic system of units. Numerous measurements of the velocity of light and of this ratio show that they are very nearly equal.

He also showed that the indices of refraction of transparent media should be equal to the square roots of their specific inductive capacities. Measurements of indices of refraction and specific inductive capacities have shown that the relation which has been stated holds true in many cases. Hopkinson has shown, however, that there are many bodies for which it does not hold true.

The theory leads to the conclusion that the direction of propagation of the electrical disturbance and the accompanying magnetic disturbance at right angles to it is normal to the plane of these disturbances. By making the assumption, which is justified by Boltzmann's measurements upon sulphur, that an eleotropic medium has different specific inductive capacities in different directions, Maxwell showed also that the propagation of the electrical disturbance in a crystal will be similar to that of light. It has also been shown that the electrical disturbance will be reflected, refracted, and polarized at a surface separating two dielectrics.

Lastly, Maxwell concluded that, if his theory be true, bodies which are transparent to the vibrations of the ether should be dielectrics, while opaque bodies should be good conductors. In the former the electrical disturbance is propagated without loss of energy; in the latter the disturbance sets up electrical currents, which heat the body, and the disturbance is not propagated through the body. Observation shows that, in fact, solid dielectrics are transparent, and solid conductors are opaque, to radiations in the

ether. Maxwell explained the fact that many electrolytes are transparent and yet are good conductors by supposing that the rapidly alternating electromotive forces which occur during the transmission of the electrical disturbance act for so short a time in one direction, that no complete separation of the molecules of the electrolyte is effected. No electrical current, therefore, is set up in the electrolyte, and electrical energy is not lost during the transmission of the disturbance.

The experiments of Hertz and others, described in § 313, have proved that electromagnetic waves may be set up in a medium, and that they possess the properties predicted for them by Maxwell's theory. In very many respects these waves behave exactly like light-waves; they are transmitted with the same velocity, they move more slowly through dense bodies than in a vacuum, they are reflected, refracted, and polarized exactly as light-waves are, and they penetrate bodies which are transparent to light, and are stopped by bodies which are opaque to light. There are certain differences between their behavior and that of light-waves, which are readily explained by the fact that the shortest electromagnetic waves which can be produced directly are several centimetres long, while none of the light-waves are as long as one one-thousandth of a millimetre. The periods of vibration of the electromagnetic waves are much greater than those of light-waves, and such properties of these waves as depend upon their periods are to some extent different from those of light.

One of the most important conclusions of the electromagnetic theory was that of the relation between the index of refraction and the specific inductive capacity. This relation is very far from being confirmed by experiment when the index of refraction is that of light. This discrepancy between theory and experiment is explained by those who maintain that light is an electromagnetic disturbance of the sort described in the following way: The methods by which the specific inductive capacity is determined either involve setting up a steady electrical force in the dielectric or the use of an alternating electrical force which at best only alternates with a

period that is enormously large in comparison with that of the light-vibrations. Since observation shows that the specific inductive capacity obtained differs when different rates of alternation are employed in the experiment, it may readily be supposed that, if alternations were used which were as rapid as those of light, values of the specific inductive capacity would be obtained which would confirm the theory. Since this cannot be done, the only method of comparison possible is to calculate, as well as it can be done, the index of refraction for light of very long period; this is done by the aid of a formula for the dispersion of light derived by Cauchy, from which the index of refraction of infinitely long waves is calculated. The agreement obtained by this method is still very far from good, but this may easily be explained by supposing that Cauchy's formula, which rests, after all, only on an empirical basis, and has been tested only within narrow limits, does not apply to waves of very long wave length, or that, in other words, the waves of long wave length exhibit anomalous dispersion. The experiments which have been made to test the relation between the specific inductive capacity and the index of refraction of electromagnetic waves show in very many cases an exceedingly good agreement, and in no case a disagreement, with the theory.

While there are still difficulties to be overcome and questions to be answered, it is yet highly probable that the true theory of light is the electromagnetic theory or some extension of it. We may therefore view magnetic, electrical, and luminous actions as actions occurring in the ether, and arising in some way from the interactions between the ether and matter, by which the energy of matter is transformed into energy in the ether, and this energy in the ether transferred through it to other matter.

TABLES.

TABLE I.

Units of Length.

Foot	=	30.48 cm.	log. 1.484015
Inch	=	2.54 cm.	log. 0.404830

Units of Mass.

Pound	=	453.59 grams.	log. 2.656664
Grain	=	0.0648 grams.	log. 8.811575

TABLE II.

Acceleration of Gravity.

$g = 980.6056 - 2.5028 \cos 2l - 0.000003h$, where l is the latitude of the station and h its height in centimetres above the sea-level.

g at Washington	=	980.07	g at Paris	=	980.94
g at New York	=	980.26	g at Greenwich	=	981.17

TABLE III.

Units of Work.

Kilogram-metre	=	$100{,}000g$ ergs.
Foot-pound	=	$13{,}825g$ ergs.
	=	1.355×10^7 ergs, log 7.13200, when $g = 980$.

Units of Rate of Working.

Watt	=	10^7 ergs per second.
Horse-power	=	550 foot-pounds per second.
	=	746 Watts.

Unit of Heat.

Lesser calorie (gram-degree) $= 4.16 \times 10^7$ ergs.

TABLE IV.

Densities of Substances at 0°.

The densities of solids given in this table must be taken as only approximate. Specimens of the same substance differ among themselves to such an extent as to render it impossible to give more precise values.

Aluminium	2.6	Iron (wrought)	7.6 to 7.8
Brass	8.4	" (cast)	7.2 to 7.7
Copper	8.9	Lead	11.3
Gold	19.3	Mercury	13.596
Glass (crown)	2.5 to 2.7	Platinum	21.5
Hydrogen	0.0000895	Silver	10.5
Ice	0.918	Zinc	7.1

TABLE V.

Units of Pressure for $g = 981$.

	Grams per sq. cm.	Dynes per sq. cm.
Pound per square inch	70.31	6.9×10^4
1 inch of mercury at 0°	34.534	3.388×10^4
1 millimetre of mercury at 0°	1.3596	1333.8
1 atmosphere (760 mm.)	1033.3	1.0136×10^6
1 atmosphere (30 inches)	1036.	1.0163×10^6

TABLE VI.

Elasticity.

If p is the force in dynes per unit area tending to extend or compress a body, the linear elasticity is $\frac{dp}{dl}$, and the volume elasticity is $\frac{dp}{dv}$.

	$\frac{dp}{dl}$.	$\frac{dp}{dv}$.
Glass	6.03×10^{11}	4.15×10^{11}
Steel	2.14×10^{12}	1.84×10^{12}
Brass	1.07×10^{12}	....
Mercury	....	3.44×10^{10}
Water	....	2.02×10^{10}

TABLE VII.

Absolute Density of Water at $t°$ in Grams per Cubic Centimetre.

$t°$.	Density.	$t°$.	Density.	$t°$.	Density.
0	0.999884	7	0.999946	40	0.99236
1	0.999941	8	0.999899	50	0.98821
2	0.999982	9	0.999837	60	0.98339
3	1.000004	10	0.999760	70	0.97795
4	1.000013	15	0.999173	80	0.97195
5	1.000003	20	0.998272	90	0.96557
6	0.999983	30	0.995778	100	0.95866

TABLE VIII.

Density of Mercury at $t°$, Water at 4° being 1.

$t°$.	Density.	log.	$t°$.	Density.	log.
0	13.5953	1.13339	20	13.5461	1.13182
10	13.5707	1.13260	30	13.5217	1.13103

TABLE IX.

Coefficients of Linear Expansion.

	Temperature.	$a = \frac{dl}{dt}$.
Aluminium	16° to 100°	0.0000235
Brass	0 to 100	0.0000188
Copper	0 to 100	0.0000167
German silver	0 to 100	0.0000184
Glass	0 to 100	0.0000071
Iron	13 to 100	0.0000123
Lead	0 to 100	0.0000280
Platinum	0 to 100	0.0000089
Silver	0 to 100	0.0000194
Zinc	0 to 100	0.0000230

Coefficients of voluminal expansion, $\frac{dV}{dt} = 3\alpha$.

TABLE X.

SPECIFIC HEATS—WATER AT 0° = 1.

Solids and Liquids.

Aluminium	0.212	Mercury	0.033
Brass	0.086	Platinum	0.032
Copper	0.093	Silver	0.056
Iron	0.112	Water (0° to 100°)	1.005
Lead	0.031	Zinc	0.056

Gases and Vapors at Constant Pressure.

Air	0.237	Nitrogen	0.244
Hydrogen	3.410	Oxygen	0.217

$$\text{Ratio, } \frac{C_p}{C_v} = 1.404.$$

TABLE XI.

I. MELTING-POINTS. II. BOILING-POINTS. III. HEATS OF LIQUEFACTION. IV. HEATS OF VAPORIZATION. V. MAXIMUM PRESSURE OF VAPOR AT 0° IN MILLIMETRES OF MERCURY.

	I.	II.	III.	IV.	V.
Ammonia	..	− 33.7°	..	294 at 7.8°	3344
Carbon dioxide	− 65°	− 78.2	..	49.3 at 0°	27100
Chlorine	..	− 33.6	..	..	4560
Copper	1200	..	..	..	..
Lead	325	..	5.9°	..	..
Mercury	− 39	357	2.8	62	0.02
Nitrous oxide, N_2O	..	− 105	..	..	24320
Platinum	1780	..	27.2	..	..
Silver	1000	..	21.1	..	..
Water	0	100	80	537	4.6
Zinc	415	..	28.1	..	..

TABLE XII.

Maximum Pressure of Vapor of Water at Various Temperatures in (I.) Dynes per Square Centimetre, (II.) Millimetres of Mercury.

Temp.	I.	II.	Temp.	I.	II.
− 20°	1236	..	60°	1.985×10^5	149.
− 10°	2790	..	80	4.729×10^5	355.
0°	6133	4.6	100	10.14×10^5	760.
10	12220	9.2	120	19.88×10^5	1491.
20	23190	17.4	140	36.26×10^5	2718.
30	42050	31.5	160	62.10×10^5	4652.
40	73200	54.6	180	100.60×10^5	7546.
50	1.226×10^5	96.2	200	$156. \times 10^5$	11689.

TABLE XIII.

Critical Temperatures (T) and Pressures in Atmospheres (P), at their Critical Temperatures, of Various Gases.

	T.	P.		T.	P.
Hydrogen	− 220.	20.	Carbon dioxide	30.9	77.
Nitrogen	− 146.	35.	Sulphur dioxide	155.4	79.
Oxygen	− 119.	51.			

TABLE XIV.

Coefficients of Conductivity for Heat (K) in C. G. S. Units, in which Q is given in Lesser Calories.

Brass	0.30	Mercury	0.015
Copper	1.11	Paraffin	0.00014
Glass	0.0005	Silver	1.09
Ice	0.0057	Vulcanized india-rubber	0.00009
Iron	0.16	Water	0.0015
Lead	0.08		

TABLE XV.

Energy Produced by Combination of 1 Gram of Certain Substances with Oxygen.

	Gram-degree of Heat.	Energy in ergs.
Carbon forming CO	2141	8.98×10^{10}
" " CO_2	8000	3.36×10^{11}
Carbon monoxide, forming CO_2	2420	1.02×10^{11}
Copper, CuO	602	2.53×10^{10}
Hydrogen, H_2O	34000	1.43×10^{12}
Marsh gas, CO_2 and H_2O	13100	5.50×10^{11}
Zinc, ZnO	1301	5.46×10^{10}

TABLE XVI.

Atomic Weights and Combining Numbers.

	Atomic Weight.		Combining Number.
Aluminium	27.04		9.01
Copper	63.18	(cupric)	31.59
"	"	(cuprous)	63.18
Gold	196.2		65.4
Hydrogen	1.		1.
Iron	55.88	(ferric)	18.63
"	"	(ferrous)	27.94
Mercury	199.8	(mercuric	99.9
"	"	(mercurous)	199.8
Nickel	58.6		29.3
Oxygen	15.96		7.98
Platinum	194.3		64.8
Silver	107.7		107.7
Zinc	64.88		32.44

TABLE XVII.

Molecular Weights and Densities of Gases.

Simple Gases.

	Atomic Weight.	Sp. gr., $H = 1$.	Mass in 1 Litre.
Chlorine, Cl_2	70.75	35.37	3.167
Hydrogen, H_2	2.00	1.00	0.0895
Nitrogen, N_2	28.024	14.012	1.254
Oxygen, O_2	31.927	15.96	1.429

Compound Gases.

	Atomic Weight.	Sp. gr., $H = 1$.	Mass in 1 Litre.
Carbonic oxide, CO..........	27.937	14.97	1.251
Carbonic dioxide, CO_2......	43.90	21.95	1.965
Hydrochloric acid, HCl......	36.376	18.188	1.628
Vapor of water, H_2O........	17.96	8.98	0.804
Atmospheric air.............			1.293

TABLE XVIII.

Electromotive Force of Voltaic Cells.

Daniell........ 1.1 volt. | Grove....... 1.88 volt. | Clark... 1.435 volt at 15°.

Electromotive force of Clark cell for any temperature t is

$$1.435[1 - 0.00077(t - 15)].$$

TABLE XIX.

Electrochemical Equivalents.

Grams per second deposited by the electromagnetic unit current,

Hydrogen, 0.0001038.

To find the electrochemical equivalents of other substances, multiply the electrochemical equivalent of hydrogen by the combining number of the substance.

TABLE XX.

Electrical Resistance.

Absolute resistance R in C. G. S. units of a centimetre cube of the substance.

Temperature coefficient, α. $R_t = R_0(1 + \alpha t)$.

	R_0.	α.
Aluminium....	2889	..
Copper..........................	1611	0.00388
German silver....	20763	0.00044
Gold............................	2041	0.00365
Iron............................	9638	..
Mercury...................... .	94340	0.00072
Platinum...........	8982	0.00376
Platinum silver, 2 Pt. 1 Ag........	24190	0.00031
Silver...........................	1580	0.00377
Zinc............................	5581	0.00365

	R_0.
Carbon (Carré's electric light)	3.9×10^{6}
Glass at 200°	2.23×10^{16}
Gutta percha, at 24°	3.46×10^{23}
" " " 0°	6.87×10^{24}
Selenium, at 100°	5.9×10^{13}
Water, at 22°	7.0×10^{10}
Zinc sulphate + $23H_2O$	1.83×10^{10}
Copper sulphate + $45H_2O$	1.91×10^{10}

TABLE XXI.

Indices of Refraction.

	Index.	Kind of Light.		Index.	Kind of Light.
Soft crown glass	1.5090	A	Canada balsam	1.528	Red
	1.5180	E	Water	1.331	B
	1.5266	G		1.336	E
Dense flint glass	1.6157	B		1.344	H
	1.6289	E	Carbon disulphide	1.614	A
	1.6453	G		1.646	E
Rock salt	1.5366	A		1.684	G
	1.5490	E	Air at 0°, 760 mm	1.00029	A
	1.5613	G		1.000296	E
Diamond	2.47	D		1.000300	H
Amber	1.532	D			

	Ordinary Index.	Kind of Light.	Extraordinary Index.
Iceland spar	1.658	D	1.486
Quartz	1.544	D	1.553

TABLE XXII.

Wave Lengths of Light—Rowland's Determinations.

Fraunhofer's line A (edge),	7593.975 tenth metres.	
B	"	6867.382
C	"	6562.965
D_1	"	5896.080
D_2	"	5890.125
E	"	5270.429
b	"	5183.735
F	"	4861.428
G	"	4307.961

TABLE XXIII.

ROTATION OF PLANE OF POLARIZATION BY A QUARTZ PLATE, 1 MM. THICK, CUT PERPENDICULAR TO AXIS.

A	12°.668	E	27°.543
B	15°.746	F	32°.773
C	17°.318	G	42°.604
D_2	21°.727	H	51°.193

TABLE XXIV.

VELOCITIES OF LIGHT.

	Cm. per Sec.		Cm. per Sec.
Michelson, 1879	2.99910×10^{10}	Foucault, 1862	2.98000×10^{10}
Michelson, 1882	2.99853×10^{10}	Cornu, 1874	2.98500×10^{10}
Newcomb, 1882	2.99860×10^{10}	Cornu, 1878	2.99990×10^{10}

THE RATIO BETWEEN THE ELECTROSTATIC AND ELECTROMAGNETIC UNITS.

	Cm. per Sec.		Cm. per Sec.
Weber and Kohlrausch	3.1074×10^{10}	Exner	2.920×10^{10}
W. Thomson	2.825×10^{10}	Klemenčič	3.018×10^{10}
Maxwell	2.88×10^{10}	Himstedt	3.007×10^{10}
Ayrton and Perry	2.98×10^{10}	Colley	3.015×10^{10}
J. J. Thomson	2.963×10^{10}		

INDEX.